Optimal k-thresholding algorithms for sparse optimization problems

Zhao, Yun-Bin

DOI:
10.1137/18M1219187

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Zhao, Y-B 2020, 'Optimal k-thresholding algorithms for sparse optimization problems', *SIAM Journal on Optimization*, vol. 30, no. 1, 25, pp. 31-55. https://doi.org/10.1137/18M1219187

SIAM J. OPTIM.
Vol. 30, No. 1, pp. 31–55

OPTIMAL k-THRESHOLDING ALGORITHMS FOR SPARSE OPTIMIZATION PROBLEMS*

YUN-BIN ZHAO†

Abstract. The simulations indicate that the existing hard thresholding technique independent of the residual function may cause a dramatic increase or numerical oscillation of the residual. This inherent drawback of the hard thresholding renders the traditional thresholding algorithms unstable and thus generally inefficient for solving practical sparse optimization problems. How to overcome this weakness and develop a truly efficient thresholding method is a fundamental question in this field. The aim of this paper is to address this question by proposing a new thresholding technique based on the notion of optimal k-thresholding. The central idea for this new development is to connect the k-thresholding directly to the residual reduction during the course of algorithms. This leads to a natural design principle for the efficient thresholding methods. Under the restricted isometry property, we prove that the optimal thresholding based algorithms are globally convergent to the solution of sparse optimization problems. The numerical experiments demonstrate that when solving sparse optimization problems, the traditional hard thresholding methods have been significantly transcended by the proposed algorithms which can even outperform the classic ℓ_1-minimization method in many situations.

Key words. sparse optimization, convex optimization, optimal k-thresholding, hard thresholding, iterative algorithms, restricted isometry property

AMS subject classifications. 90C25, 90C05, 90C30, 65F10, 94A12, 15A29

DOI. 10.1137/18M1219187

1. Introduction.

Let $A \in \mathbb{R}^{m\times n}$ $(m < n)$ be a given matrix, $y \in \mathbb{R}^m$ be a given vector, and $\varepsilon \geq 0$ be a given parameter. Let $\|x\|_0$ denote the "ℓ_0-norm" counting the number of nonzero entries of the vector $x \in \mathbb{R}^n$. The sparse optimization problem is to find a sparse (or the sparsest) vector, denoted by x^*, such that Ax^* can best fit the vector y. This problem can be formulated as the minimization problem with a sparsity constraint

$$\min_x\{\|Ax - y\|_2^2 : \|x\|_0 \leq k\}, \tag{1.1}$$

where k is a prescribed integer number, or formulated as the so-called ℓ_0-minimization problem

$$\min_x\{\|x\|_0 : \|Ax - y\|_2 \leq \varepsilon\}. \tag{1.2}$$

Both (1.1) and (1.2) are the central models for sparse signal recovery and sparse representation of data on their redundant bases. These models provide an essential basis for the development of the theory and algorithms for compressed sensing (see, e.g., [12, 25, 26, 32, 53]). The problem (1.1) has also been widely used in the fields of statistical regressions and wireless communications (see, e.g., [45, 4, 41]).

The problems (1.1) and (1.2) are NP-hard in general [46]. The plausible algorithms for such problems can be briefly categorized into the following classes: (i) convex optimization methods (e.g., ℓ_1-minimization [19], reweighed ℓ_1-minimization

*Received by the editors October 5, 2018; accepted for publication (in revised form) October 22, 2019; published electronically January 2, 2020.
https://doi.org/10.1137/18M1219187
†School of Mathematics, The University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK (y.zhao.2@bham.ac.uk).

[16, 31, 56], and dual-density-based reweighted ℓ_1-minimization [53, 54, 55]); (ii) heuristic methods (such as matching pursuit [43], orthogonal matching pursuit [44, 52], compressive sampling matching pursuit [47], and subspace pursuit [20]); (iii) thresholding methods (e.g., soft thresholding [21, 22, 24], hard thresholding [6, 7, 8, 30], graded hard thresholding pursuits [10, 11], and the "firm" thresholding [51]); (iv) integer programming methods [4].

The use of thresholding techniques for signal denoising problems can be dated back to the seminal paper by Donoho and Johnstone [23]. Since then, various thresholding algorithms were proposed for sparse recovery or sparse approximation (see, e.g., Reeves and Kingsbury [49], Kingsbury and Reeves [39], Figueiredo and Nowak [27], Starck, Nguyen, and Murtagh [50], Herrity, Gilbert, and Tropp [35], Blumensath and Davies [7, 8, 9], and Beck and Teboulle [3]). The thresholding algorithms can be derived from different perspectives such as minimizing certain surrogate functions related to the residual function $\|y - Ax\|_2^2$ (see, e.g., [7, 21, 37]) and the necessary optimality conditions for minimization with sparsity constraints [1, 2]. The algorithms can be classified as soft thresholdings or hard thresholdings according to the nature of thresholding operators. The soft ones are closely related to the optimality condition of certain convex optimization (e.g., [24, 51]) and have been widely analyzed in the literature (e.g., [21, 24, 28, 35, 51]). The hard thresholding ones for compressed sensing were analyzed by Blumensath and Davies [7, 8, 9], Foucart [29, 30], and Foucart and Rauhut [32].

For convenience of discussion, we focus on the problem (1.1) in this paper. Given $z \in \mathbb{R}^n$, let $\mathcal{H}_k(z)$ denote the vector obtained by retaining the k largest magnitudes of z and zeroing out the remaining entries of z. The operator $\mathcal{H}_k(\cdot)$ is referred to as the *hard thresholding operator*. Since the k largest magnitudes of z may not be unique (see Theorem 2.3 for details), $\mathcal{H}_k(z)$ might contain more than one vector in some situations. The iterative hard thresholding (IHT) algorithm takes the scheme

$$x^{p+1} \in \mathcal{H}_k\left(x^p + \tau A^T(y - Ax^p)\right) \tag{1.3}$$

to search the solution of (1.1), where A^T is the transpose of A and $\tau > 0$ is a stepsize which can be iteratively updated or a fixed number (such as $\tau \equiv 1$). The iterative scheme (1.3) can be dated back to Landweber [36]. The Landweber iteration $z^{p+1} = z^p + \tau A^T(y - Az^p)$ is essentially the gradient method for minimizing the function $\|y - Ax\|_2^2$. Thus an intuitive idea for possibly solving the problem (1.1) is to perform the hard thresholding on the Landweber iteration, leading to the iterative scheme (1.3).

The analyses in [7, 8, 29, 30, 42] show that the convergence of the IHT algorithm can be guaranteed under the restricted isometry property (RIP) or a mutual coherence condition. The RIP was first introduced by Candès and Tao [15] (see also Candès [14]) to study the signal recovery via the ℓ_1-minimization method. However, the empirical evidences indicate that the efficiency of the IHT is actually low. For instance, taking $A = \left[\begin{smallmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{smallmatrix}\right]$ and $y = \left[\begin{smallmatrix} 1 \\ 5 \end{smallmatrix}\right]$, it is evident that $x^* = (1, 0, 0, 0)^T$ is the solution to the problem (1.1). However, the IHT starting from $x^0 = 0$ generates the following sequence: $x^1 = \mathcal{H}_1(u^0) = (0, 0, 0, 44)$, $x^2 = \mathcal{H}_1(u^1) = (0, 0, 0, -3432)$, $x^3 = \mathcal{H}_1(u^2) = (0, 0, 0, -271170), \ldots$, where $u^p := x^p + A^T(y - Ax^p)$. The sequence $\{x^p\}$ diverges, and the corresponding sequence of $r(x^p) = \|y - Ax^p\|_2$ (i.e., $r(x^0) = \sqrt{26}$, $r(x^1) = 388.6309$, $r(x^2) = 3.0702e{+}04$, $r(x^3) = 2.4254e{+}06, \ldots$) also diverges so quickly. Thus there is a huge gap between the theoretical efficiency

and practical performance of the IHT. This stimulates the study of various acceleration and stabilization techniques for this sort of algorithm.

The first idea for acceleration is using a stepsize as in (1.3). The algorithm with a fixed stepsize was called gradient descent with sparsification in [33]. See also [1, 6, 17]. With iteratively updated stepsizes, the algorithm is called the normalized IHT in [9]. See also the so-called conjugate gradient IHT algorithm in [5]. Another idea is to minimize the residual over the support determined by the hard thresholding. With this idea, Foucart [30] proposed the following algorithm called hard thresholding pursuit (HTP):

$$S^{p+1} = \mathrm{supp}(\hat{z}),\ \hat{z} \in \mathcal{H}_k(x^p + A^T(y - Ax^p)), \tag{1.4}$$

$$x^{p+1} \in \arg\min_x \left\{ \|y - Ax\|_2^2 :\ \mathrm{supp}(x) \subseteq S^{p+1} \right\}. \tag{1.5}$$

The step (1.5) is used to chase a better vector than $\hat{z}$ that can best fit the vector y. This idea is also used in compressive sensing matching pursuit proposed by Needell and Tropp [47] and in subspace pursuit proposed by Dai and Milenkovic [20]. As a generalization of the HTP, the graded HTP [10, 11] combines the step (1.5) and orthogonal matching pursuit. Other acceleration versions of the IHT based on Nesterov's techniques [48] can be found in [17, 38, 40].

In many situations, however, directly using the operator $\mathcal{H}_k(\cdot)$ is not attractive from the perspective of the residual $\|y - Ax\|_2^2$. The thresholding step (1.4) is actually independent of the residual reduction (see section 3 for details) in the sense that it does not include any mechanism to reduce the residual in the course of iterations. It actually causes the divergence of the IHT in numerous situations, or significantly slows down the convergence of the algorithm. Even aided with (1.5), numerical experiments demonstrate that the values of the residual at the iterates generated by the HTP may still oscillate dramatically, rendering the algorithm inefficient in many situations. Such an oscillation phenomenon (see Figure 5.1(a) in section 5) was caused by the hard thresholding operator which often increases instead of decreasing the residual. To our knowledge, the existing ideas for acceleration do not serve the purpose of eliminating such an inherent drawback of the operator $\mathcal{H}_k$.

In this paper, retaining k entries of a vector and zeroing out its remaining entries is referred to as a k-thresholding of the vector. Motivated by the above observation, we explore the following idea in order to develop efficient thresholding methods: *The k-thresholding should be performed to serve the purpose of residual reduction.* Linking the thresholding with residual reduction enables us to introduce the notion of *optimal k-thresholding*. More specifically, it enables us to select a set of k entries of a vector that achieves the least residual among all possible selections of k entries. Clearly, the optimal k entries is not necessarily the k largest magnitudes of the vector. Based on this notion, we propose the optimal k-thresholding (OT) algorithm and the optimal k-thresholding pursuit (OTP). Since the subproblems in OT and OTP are binary quadratic minimization problems which are usually not convenient to solve directly, we propose the relaxed optimal k-thresholding (ROT) and the relaxed optimal k-thresholding pursuit (ROTP) which naturally result from the tightest convex relaxation of the binary optimization problem in OT and OTP. The ROTP and its further enhanced versions (ROTP2 and ROTP3) turn out to be a new and powerful generation of thresholding algorithms which significantly reverse the adversity of using the traditional hard thresholding.

The OT and OTP algorithms are shown to have the guaranteed success for sparse signal recovery under the RIP bound $\delta_{2k} < 0.5349$ (see Theorem 4.3 for details). This bound is largely theoretical by assuming that the binary subproblems in OT or OTP can be successfully solved by certain methods. The guaranteed success of the ROT and ROTP is also proved in this paper under the RIP bound $\delta_{3k} < 1/5$. The empirical results collected from random examples of sparse optimization problems show that the ROTP and its enhanced versions remarkably outperform the IHT and HTP as anticipated, and the ROTP2 and ROTP3 are efficient enough to outperform the ℓ_1-minimization in numerous situations (see section 5 for details). Simulations also demonstrate that the proposed algorithms are stable in the sense that the residual is steadily reduced during the course of iterations.

The paper is organized as follows. Section 2 provides some notations, definitions, and properties of the hard thresholding operator. The new thresholding methods are described in section 3. The theoretical performance of several proposed algorithms is rigorously shown under the RIP condition in section 4. Numerical results for the ROTP and its enhanced versions are reported in section 5.

2. Preliminary.

2.1. Notation. $\mathbb{R}^n$ denotes the n-dimensional Euclidean space, and $\mathbb{R}^{m\times n}$ stands for the set of $m \times n$ metrics. The set of n-dimensional binary vectors is denoted by $\{0,1\}^n$. All vectors are column vectors unless otherwise specified. We use $\mathbf{e}$ to denote the vector of ones and I to denote the identity matrix. For a vector $x \in \mathbb{R}^n$, $\|x\|_2$, $\|x\|_1$, and $\|x\|_\infty$ denote the ℓ_2-, ℓ_1-, and ℓ_∞-norms, respectively, and $|x|$ denotes the absolute vector of x, i.e., $|x|_i = |x_i|$ for $i = 1,\dots,n$. The support of x is denoted by $\mathrm{supp}(x)$ which is the index set $\{i : x_i \neq 0\}$. The nonnegative vector x is written as $x \geq 0$. For two vectors x and y, the inequality $x \geq y$ means $x - y$ is a nonnegative vector. Given a set $S \subseteq \{1,2,\dots,n\}$, $|S|$ denotes the cardinality of S, and $\overline{S} = \{1,2,\dots,n\}\backslash S$ denotes the complement set of S. Given $x \in \mathbb{R}^n$ and $S \subseteq \{1,\dots,n\}$, $x_S \in \mathbb{R}^n$ denotes the vector obtained by retaining the components of x indexed by S and zeroing out the remaining components of x. That is, for every $i = 1,\dots,n$, $(x_S)_i = x_i$ if $i \in S$; otherwise, $(x_S)_i = 0$. For $x, y \in \mathbb{R}^n$, the vector $x \otimes y$ is the Hadamard product of x and y, i.e., $x \otimes y = (x_1y_1,\dots,x_ny_n)^T$. A vector is said to be k-sparse if $\|x\|_0 \leq k$.

2.2. Characteristics of hard thresholding operator $\mathcal{H}_k(\cdot)$. Given an integer number k, a vector $w \in \{0,1\}^n$ with exactly k nonzero entries can be represented as $w \in \{0,1\}^n$ and $\mathbf{e}^T w = k$. Denote the set of such vectors by

$$\mathcal{W}^{(k)} = \left\{w : \ w \in \{0,1\}^n, \ \mathbf{e}^T w = k\right\}. \tag{2.1}$$

Note that the Hadamard product $\widetilde{z} = z \otimes w$, where $z \in \mathbb{R}^n$ and $w \in \mathcal{W}^{(k)}$, is the vector thresholded from z by retaining z_i corresponding to $w_i = 1$ and zeroing out the remaining ones. We introduce the following definition.

DEFINITION 2.1. *Given $z \in \mathbb{R}^n$ and $w \in \mathcal{W}^{(k)}$, the vector $\widetilde{z} = z \otimes w$ is called a k-thresholding vector of z, and the associated vector $w \in \mathcal{W}^{(k)}$ is called a k-thresholding indicator. If the k-thresholding retains the k largest magnitudes of z, it is referred to as the hard k-thresholding of z.*

Clearly, $\mathcal{H}_k(z)$ is the set of $z \otimes w^\#$, where $w^\# \in \mathcal{W}^{(k)}$ is an indicator for the k largest magnitudes of z. Denote by $\mathcal{W}^*(z) \subseteq \mathcal{W}^{(k)}$ the set of indicators for the k largest magnitudes of z. Then $\mathcal{W}^*(z) = \{w^\# \in \mathcal{W}^{(k)} : \ z \otimes w^\# \in \mathcal{H}_k(z)\}$. We also

note that $\mathcal{H}_k(z)$ is the set of vectors which are the best k-term approximation of z, namely,

$$\mathcal{H}_k(z) = \arg\min_u \{\|z-u\|_1 : \ \|u\|_0 \le k\}.$$

Denote by $\sigma_k(z)_1$ the error of the best k-term approximation of z, i.e.,

$$\sigma_k(z)_1 = \min_u \{\|z-u\|_1 : \ \|u\|_0 \le k\}.$$

Clearly, $\sigma_k(z)_1 = 0$ if and only if z is k-sparse. In this paper, z is said to be k-compressible if $\sigma_k(z)_1$ is small. Note that $\sigma_k(z)_1 = \|z - \hat{z}\|_1$ for any $\hat{z} \in \mathcal{H}_k(z)$. By the definition of $\mathcal{W}^*(z)$ and $\mathcal{W}^{(k)}$, we see that for any $w^\# \in \mathcal{W}^*(z)$

$$\sigma_k(z)_1 = \|z - (z \otimes w^\#)\|_1 = \|z \otimes (\mathbf{e} - w^\#)\|_1 = |z|^T(\mathbf{e} - w^\#). \tag{2.2}$$

Since $\|z \otimes w\|_0 \le k$ for any $w \in \mathcal{W}^{(k)}$, by the definition of $\sigma_k(z)_1$, we have

$$\sigma_k(z)_1 \le \|z - (z \otimes w)\|_1 = |z|^T(\mathbf{e} - w) \ \text{ for any } w \in \mathcal{W}^{(k)}. \tag{2.3}$$

It follows from (2.2) and (2.3) that every hard k-thresholding indicator $w^\# \in \mathcal{W}^*(z)$ is exactly the solution to the following 0-1 integer programming problem:

$$\min_w \left\{ |z|^T(\mathbf{e} - w) : \ \mathbf{e}^T w = k, \ w \in \{0,1\}^n \right\}. \tag{2.4}$$

This problem is very easy to solve via the linear programming (LP) relaxation

$$\min_w \left\{ |z|^T(\mathbf{e} - w) : \ \mathbf{e}^T w = k, \ 0 \le w \le \mathbf{e} \right\}, \tag{2.5}$$

as indicated by the following lemma.

LEMMA 2.2. *Given $z \in \mathbb{R}^n$, let $\hat{\gamma}(z)$ be the optimal objective value of* (2.5)*, and let $\widehat{S}$ be the set of optimal solutions of* (2.5) *that are extreme points of the feasible set. Then $\hat{\gamma}(z) = \sigma_k(z)_1$ and $\widehat{S} = \mathcal{W}^*(z)$. Thus $w^\# \in \widehat{S}$ if and only if $z \otimes w^\# \in \mathcal{H}_k(z)$.*

Proof. Consider the feasible set of the problem (2.5)

$$\mathcal{P} := \{w \in \mathbb{R}^n : \ \mathbf{e}^T w = k, \ 0 \le w \le \mathbf{e}\}. \tag{2.6}$$

Let V denote the set of extreme points of this polyhedron. By introducing the nonnegative variable $u \in \mathbb{R}^n$, the linear system in $\mathcal{P}$ can be written as $\mathbf{e}^T w = k$, $w + u = \mathbf{e}$, $w \ge 0$ and $u \ge 0$, that is, $\begin{bmatrix} \mathbf{e}^T & 0 \\ I & I \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix} = \begin{bmatrix} k \\ \mathbf{e} \end{bmatrix}$ and $\begin{bmatrix} w \\ u \end{bmatrix} \ge 0$, where I is the $n \times n$ identity matrix. Note that the matrix $\begin{bmatrix} \mathbf{e}^T & 0 \\ I & I \end{bmatrix}$ is totally unimodular, and the right-hand-side vector $\begin{bmatrix} k \\ \mathbf{e} \end{bmatrix}$ of the above system is an integer vector. The total-unimodularity theory implies that every extreme point of the polyhedron $\mathcal{P}$ is an integer vector. Therefore, by the structure of $\mathcal{P}$, every extreme point of $\mathcal{P}$ must be a binary vector with k entries being ones. This means $V \subseteq \mathcal{W}^{(k)}$. By the LP theory, at least one of the extreme points of $\mathcal{P}$ must be optimal. Thus $\emptyset \ne \widehat{S} \subseteq V \subseteq \mathcal{W}^{(k)}$. It follows from (2.3) that $\sigma_k(z)_1 \le \hat{\gamma}(z) = |z|^T(\mathbf{e} - \hat{w})$ for $\hat{w} \in \widehat{S} \subseteq \mathcal{W}^{(k)}$. Let $\tilde{w} \in \mathcal{W}^*(z)$ which is contained in the feasible set of (2.5). By optimality and (2.2), we have $\hat{\gamma}(z) \le |z|^T(\mathbf{e} - \tilde{w}) = \sigma_k(z)_1$. Therefore $\sigma_k(z)_1 = \hat{\gamma}(z)$, from which it is not difficult to see that $\mathcal{W}^*(z)$ is exactly the set of optimal solutions of (2.5) that are extreme points of the feasible set, and hence $\widehat{S} = \mathcal{W}^*(z)$. Therefore $w^\# \in \widehat{S}$ if and only if $z \otimes w^\# \in \mathcal{H}_k(z)$. ∎

It is well known that finding the hard k-thresholding of a vector is very easy and can be done in several ways. The equivalence of (2.4) and (2.5) implies that solving the LP problem (2.5) is an alternative way. We now point out that the condition for $\mathcal{H}_k(z)$ being a singleton can be completely characterized. Denote by z^* the nonincreasing rearrangement of $|z|$, i.e., $z_1^* \geq z_2^* \geq \cdots \geq z_n^* \geq 0$, and π is a permutation of $\{1, \ldots, n\}$ such that $z_j^* = |z_{\pi(j)}|$ for $j = 1, \ldots, n$. The following theorem claims that $\mathcal{H}_k(z)$ is a singleton if and only if the kth largest absolute entry of z is strictly larger than the $(k+1)$th largest absolute entry.

THEOREM 2.3. *Let $z \in \mathbb{R}^n$ be a given vector and z^* be the nonincreasing arrangement of $|z|$. The following three statements are equivalent:* (a) *$\mathcal{H}_k(z)$ is a singleton;* (b) *the solution of the LP problem* (2.5) *is unique;* (c) *$z_k^* > z_{k+1}^*$.*

Proof. The equivalence of (a) and (b) follows from Lemma 2.2 straightaway. It is sufficient to show the equivalence of (c) and (a). First we note that when $z_k^* = z_{k+1}^*$, there are at least two distinct sets of the k largest magnitudes of z, so $\mathcal{H}_k(z)$ is not unique. Thus (a) implies (c). We now show that (c) also implies (a). Assume that $z_k^* > z_{k+1}^*$ and denote by the set $L_k(z) = \{\pi(i) : \ |z_{\pi(i)}| = z_i^*, \ i = 1, \ldots, k\}$, which is the set of indices for the k largest magnitudes of z. Let w^* be an arbitrary optimal solution of (2.5) which is an extreme point of its feasible set. Note that

$$\sum_{i \notin L_k(z)} |z_i| w_i^* \leq \left[\max_{i \notin L_k(z)} |z_i| \right] \sum_{i \notin L_k(z)} w_i^* = z_{k+1}^* \left[k - \sum_{i \in L_k(z)} w_i^* \right] = z_{k+1}^* \sum_{i \in L_k(z)} (1 - w_i^*).$$

From Lemma 2.2, we have $\sigma_k(z)_1 = |z|^T(\mathbf{e} - w^*)$. This together with the inequality above implies

$$\begin{aligned}
\sigma_k(z)_1 &= \sum_{i \in L_k(z)} |z_i|(1 - w_i^*) + \sum_{i \notin L_k(z)} |z_i|(1 - w_i^*) \\
&= \sum_{i \notin L_k(z)} |z_i| + \sum_{i \in L_k(z)} |z_i|(1 - w_i^*) - \sum_{i \notin L_k(z)} |z_i| w_i^* \\
&\geq \sigma_k(z)_1 + \sum_{i \in L_k(z)} |z_i|(1 - w_i^*) - z_{k+1}^* \sum_{i \in L_k(z)} (1 - w_i^*) \\
&= \sigma_k(z)_1 + \sum_{i \in L_k(z)} (|z_i| - z_{k+1}^*)(1 - w_i^*).
\end{aligned}$$

With the fact $0 \leq w^* \leq \mathbf{e}$ and $|z_i| > z_{k+1}^*$ for every $i \in L_k(z)$, the inequality above implies that $w_i^* = 1$ for all $i \in L_k(z)$. By the constraints of (2.5), the remaining $n - k$ components of w^* are equal to 0. So w^* is uniquely determined. This means the set of the optimal solutions of (2.5) which are extreme points of its feasible set contains only a single vector, and thus $\mathcal{H}_k(z)$ is unique (by Lemma 2.2). ∎

The link between $\mathcal{H}_k$, $\mathcal{P}$, and $\sigma_k(\cdot)_1$ indicates that performing $\mathcal{H}_k(\cdot)$ on a vector is nothing but minimizing the error of the k-term approximation of the vector, which is independent of the residual function $\|y - Az\|_2^2$. This motivates us to consider a new thresholding strategy in the next section.

3. Optimal *k*-thresholding algorithms and their relaxations. The classic steepest descent method for minimizing the residual $\|y - Ax\|_2^2$ is deeply rooted in the following theoretical basis: When the current iterate is not a minimizer of the function, moving from the iterate in the direction of negative gradient of the function

(with a certain stepsize if necessary) leads to the decrease in the value of this function. This theoretical basis, however, is generally lost when the operator $\mathcal{H}_k(\cdot)$ is applied to the vector $u^p := x^p + A^T(y - Ax^p)$. As we have pointed out in section 2, the selection of the k largest magnitudes of this vector is independent of the residual $\|y - Ax\|_2^2$. Thus the hard k-thresholding may cause the increase of the residual at $\hat{u} \in \mathcal{H}_k(u^p)$, i.e., $\|y - A\hat{u}\|_2 > \|y - Ax^p\|_2$. This is the main reason for the iterative scheme $x^{p+1} \in \mathcal{H}_k(u^p)$ being unstable and inefficient for solving sparse optimization problems, unless u^p is k-compressible (in which case $\mathcal{H}_k(u^p) \approx u^p$) so that the scheme $x^{p+1} \in \mathcal{H}_k(u^p)$ is close to the steepest descent method.

To overcome the drawback of the hard thresholding, we may link the k-thresholding with a residual function and perform thresholding and residual reduction simultaneously. This stimulates the following thresholding of a vector $z \in \mathbb{R}^n$:

$$\alpha^*(u) := \min_w \{\|y - A(u \otimes w)\|_2^2 : \ \mathbf{e}^T w = k, \ w \in \{0,1\}^n\}. \tag{3.1}$$

In this model, performing a k-thresholding of u is directly related to the residual function. The k-thresholding of u resulting from (3.1) admits the least residual, and thus it is better than other k-thresholdings of u, including $\mathcal{H}_k(u)$. We use $w^*(u)$ to denote the optimal solution of (3.1). Clearly, the solution $w^*(u)$ relies on the choice of the objective function, which may take other forms different from the one in (3.1). For instance, we may minimize the ℓ_1-norm of the gradient of $\|y - Ax\|_2^2$, leading to the following model:

$$\min_w \left\{ \|A^T(y - A(u \otimes w))\|_1 : \ \mathbf{e}^T w = k, \ w \in \{0,1\}^n \right\}.$$

For simplicity, however, we only focus on the model (3.1) and its convex relaxations in this paper. We introduce the following definition.

DEFINITION 3.1. *Given $u \in \mathbb{R}^n$, the solution of* (3.1)*, denoted by $w^*(u)$, is called the optimal k-thresholding indicator, and the vector $u \otimes w^*(u)$ is called the optimal k-thresholding of u. The operator*

$$Z_k^{\#}(u) := \{u \otimes w^*(u) : \ w^*(u) \text{ is an optimal solution of } (3.1)\}$$

is called the optimal k-thresholding operator.

The solution of (3.1) may not be unique, and thus $Z_k^{\#}(u)$ might contain more than one vector. Since $\alpha^*(u) = \|y - Av\|_2^2$ for any $v \in Z_k^{\#}(u)$, we may simply write this as $\alpha^*(u) = \|y - AZ_k^{\#}(u)\|_2^2$ no matter if $Z_k^{\#}(u)$ is a singleton or not. By optimality, we have

$$\left\|y - AZ_k^{\#}(u)\right\|_2 \le \|y - A(u \otimes w)\|_2 \ \text{ for any } w \in \mathcal{W}^{(k)}, \tag{3.2}$$

where $\mathcal{W}^{(k)}$ is given in (2.1). This implies that

$$\left\|y - AZ_k^{\#}(u)\right\|_2 \le \min_{\hat{u} \in \mathcal{H}_k(u)} \|y - A\hat{u}\|_2. \tag{3.3}$$

Thus the optimal k-thresholding is never worse than the hard k-thresholding from the perspective of residual reduction. In terms of optimal k-thresholding, we obtain the following iterative scheme:

$$x^{p+1} \in Z_k^{\#}\left(x^p + A^T(y - Ax^p)\right). \tag{3.4}$$

This method is referred to as the optimal k-thresholding (OT) algorithm, which by the definition of $Z_k^{\#}(\cdot)$ is described explicitly as follows.

OT algorithm. Input (A, y, k). Give an initial point $x^0 \in \mathbb{R}^n$ and repeat the following steps until a stoping criterion is satisfied:

S1. At x^p, set $u^p = x^p + A^T(y - Ax^p)$ and solve the problem

$$\min_w \left\{ \|y - A(u^p \otimes w)\|_2^2 : \ \mathbf{e}^T w = k, \ w \in \{0,1\}^n \right\}. \tag{3.5}$$

S2. Let $w^*(u^p)$ be the solution to the problem (3.5), and set

$$x^{p+1} := u^p \otimes w^*(u^p).$$

In general, the input vector u^p in S1 is not k-sparse, but the output $u^p \otimes w^*(u^p)$ of S1 is a compressed (in fact, k-sparse) vector. So the step S1 above can be called a "compressing step." The binary optimization problem (3.5) is known to be NP-hard [18] (see also [13]). This problem is similar to the best subset selection model in statistics [45], and Bertsimas, King, and Mazumder [4] developed a mixed-integer optimization formulation to deal with similar binary optimization problems. Their study indicates that in many cases the problem like (3.5) can be directly solved by exploiting the integer programming structure, and thus it might not be always necessary to consider a convex relaxation of the problem (see the numerical results in [4] for more details).

In this paper, however, we focus on the convex relaxation of the binary problem (3.5). The convex relaxation turns out to be a very efficient technique for the development of practical thresholding algorithms based on the above OT framework. To relax the problem (3.5), an immediate idea is to replace the binary constraint $w \in \{0,1\}^n$ with the simple restriction $w \in [0,1]^n$. In other words, we replace the feasible set $\mathcal{W}^{(k)}$ of (3.5) with the polytope $\mathcal{P}$ defined in (2.6). From the proof of Lemma 2.2, we see that $\mathcal{P}$ is the convex hull, i.e., the tightest convex relaxation of $\mathcal{W}^{(k)}$. This leads to the following convex relaxation counterpart of (3.5):

$$\gamma^*(u^p) := \min_w \left\{ \|y - A(u^p \otimes w)\|_2^2 : \ \mathbf{e}^T w = k, \ 0 \le w \le \mathbf{e} \right\}, \tag{3.6}$$

which can be solved efficiently by interior-point methods or other optimization methods. Let w^p be the solution of (3.6). Since w^p may not be exactly k-sparse, we apply $\mathcal{H}_k$ to the vector $u^p \otimes w^p$ to produce the next k-sparse iterate. This leads to following relaxed optimal k-thresholding method termed the "ROT" algorithm.

ROT algorithm. Input (A, y, k). Give an initial point x^0 and repeat the following steps until a stoping criterion is satisfied:

S1. At x^p, set $u^p = x^p + A^T(y - Ax^p)$ and solve the convex optimization problem (3.6) to obtain w^p.

S2. Set

$$x^{p+1} \in \mathcal{H}_k(u^p \otimes w^p).$$

The first step above can still be called a "compressing step" since the output $u^p \otimes w^p$ is more compressible than u^p in the sense that $\sigma_k(u^p \otimes w^p)_1 \le \sigma_k(u^p)_1$ which follows from the fact $0 \le w^p \le \mathbf{e}$. In fact, for a given vector $z \in \mathbb{R}^n$, $\sigma_k(z)_1$ is the sum of the $n-k$ smallest components of $|z|$. Let Λ denote the index set of the $n-k$ smallest components of $|u^p|$. Then $\sigma_k(u^p)_1 = \|(u^p)_\Lambda\|_1$. Therefore,

$$\sigma_k(u^p \otimes w^p)_1 \le \|(u^p \otimes w^p)_\Lambda\|_1 \le \|(w^p)_\Lambda\|_\infty \|(u^p)_\Lambda\|_1 \le \sigma_k(u^p)_1,$$

where the last inequality follows from the fact $\|(w^p)_\Lambda\|_\infty \le 1$. When $k \ll n$ (which is typical in compressed sensing scenarios), most components of w^p are very small, and thus $\sigma_k(u^p \otimes w^p)_1$ might be much smaller than $\sigma_k(u^p)_1$. In particular, $\sigma_k(u^p \otimes w^p)_1 = 0$ when $w^p \in \mathcal{W}^{(k)}$. The difference between the ROT and traditional hard thresholding methods is obvious. Traditional ones directly apply the hard k-thresholding to u^p without making any effort to reduce the residual. When u^p is not compressible, the hard thresholding $\mathcal{H}_k(u^p)$ might dramatically raise the value of the residual function, causing divergence or very slow convergence of the iterates. By contrast, the ROT improves the efficiency of thresholdings by simultaneously compressing the vector u^p and decreasing the residual. The ROT integrates these two efforts to overcome the drawback of performing $\mathcal{H}_k(\cdot)$ directly onto noncompressible vectors. We now point out an advantage of applying $\mathcal{H}_k(\cdot)$ to a compressible vector.

LEMMA 3.1. *Let u be an arbitrary vector in $\mathbb{R}^n$. Then for any $\hat{u} \in \mathcal{H}_k(u)$,*

$$\big|\|y - A\hat{u}\|_2^2 - \|y - Au\|_2^2\big| \le 2\|A^T(y - Au)\|_\infty \sigma_k(u)_1 + \lambda_{\max}(A^TA)(\sigma_k(u)_1)^2. \tag{3.7}$$

Proof. Let $\hat{u} \in \mathcal{H}_k(u)$. Note that

$$\|y - A\hat{u}\|_2^2 = \|y - Au\|_2^2 + 2[A^T(y - Au)]^T(\hat{u} - u) + (\hat{u} - u)^T A^T A(\hat{u} - u).$$

Thus,

$$\big|\|y - A\hat{u}\|_2^2 - \|y - Au\|_2^2\big| \le 2\|A^T(y - Au)\|_\infty \|\hat{u} - u\|_1 + \lambda_{\max}(A^TA)\|\hat{u} - u\|_2^2,$$

which together with $\|\hat{u} - u\|_2 \le \|\hat{u} - u\|_1 = \sigma_k(u)_1$ implies the inequality (3.7). ∎

This lemma shows that if $\sigma_k(u)_1$ is small (i.e., u is k-compressible), then $\|y - A\hat{u}\|_2^2 \approx \|y - Au\|_2^2$ for any $\hat{u} \in \mathcal{H}_k(u)$. Thus performing a hard k-thresholding on a compressible vector will not dramatically raise the value of the residual. Since the output, $u^p \otimes w^p$, of the first step of ROT is more compressible than the input vector u^p, the way for generating x^{p+1} in ROT is believed to be more sensible than the way in IHT and HTP. The next result interprets further why a hard k-thresholding should apply to compressible vectors instead of noncompressible ones.

THEOREM 3.2. *Let $x^p \in \mathbb{R}^n$ be given and $u^p = x^p + A^T(y - Ax^p)$. Let $\gamma^*(u^p)$ and w^p be the optimal value and the optimal solution of* (3.6)*, respectively, and let $x^{p+1} \in \mathcal{H}_k(u^p \otimes w^p)$. Denote by $\alpha^*(u^p)$ the optimal value of* (3.5)*. Then the following two statements hold:* (i) $\gamma^*(u^p) \le \alpha^*(u^p) \le \min_{\hat{u} \in \mathcal{H}_k(u^p)} \|y - A\hat{u}\|_2^2$; (ii) $\|y - Ax^{p+1}\|_2^2 \le \alpha^*(u^p)$ *provided that*

$$\sigma_k(u^p \otimes w^p)_1 \le \frac{\sqrt{\varphi(u^p, w^p)^2 + 4(\alpha^*(u^p) - \gamma^*(u^p))\lambda_{\max}(A^TA)} - \varphi(u^p, w^p)}{2\lambda_{\max}(A^TA)},$$

where $\varphi(u^p, w^p) = 2\|A^T(y - A(u^p \otimes w^p))\|_\infty$.

Proof. The statement (i) is obvious, following directly from (3.3) and the optimality of w^p. Let $\varphi(u^p, w^p)$ be defined as above. Consider the following quadratic function (in variable t): $\vartheta(t) = \gamma^*(u^p) + \varphi(u^p, w^p)t + \lambda_{\max}(A^TA)t^2$. By Lemma 3.1,

$$\begin{aligned}
\|y - Ax^{p+1}\|_2^2 &\le \|y - A(u^p \otimes w^p)\|_2^2 + 2\|A^T(y - A(u^p \otimes w^p))\|_\infty \sigma_k(u^p \otimes w^p)_1 \\
&\quad + \lambda_{\max}(A^TA)(\sigma_k(u^p \otimes w^p)_1)^2 \\
&= \gamma^*(u^p) + \varphi(u^p, w^p)\sigma_k(u^p \otimes w^p)_1 + \lambda_{\max}(A^TA)(\sigma_k(u^p \otimes w^p)_1)^2 \\
&= \vartheta(\sigma_k(u^p \otimes w^p)_1). \qquad (3.8)
\end{aligned}$$

It is easy to verify that $\vartheta(t) \le \alpha^*(u^p)$ provided that t is smaller than or equal to the following root of the quadratic equation $\vartheta(t) = \alpha^*(u^p)$:

$$\Omega(u^p, w^p) := \frac{-\varphi(u^p, w^p) + \sqrt{\varphi(u^p, w^p)^2 + 4(\alpha^*(u^p) - \gamma^*(u^p))\lambda_{\max}(A^TA)}}{2\lambda_{\max}(A^TA)}.$$

Thus when $\sigma_k(u^p \otimes w^p)_1 \le \Omega(u^p, w^p)$, we must have $\vartheta(\sigma_k(u^p \otimes w^p)_1) \le \alpha^*(u^p)$. This, combined with (3.8), implies that $\|y - Ax^{p+1}\|_2^2 \le \alpha^*(u^p)$. ∎

This result shows that if $\sigma_k(u^p \otimes w^p)_1$ is small enough, then

$$\max_{\bar{u} \in \mathcal{H}_k(u^p \otimes w^p)} \|y - A\bar{u}\|_2 \le \|y - AZ_k^{\#}(u^p)\|_2 \le \min_{\hat{u} \in \mathcal{H}_k(u^p)} \|y - A\hat{u}\|_2,$$

which means the iterates generated by the ROT will never be worse than the traditional hard thresholding algorithms from the perspective of residual reductions. The OT and ROT algorithms can be further enhanced by using the pursuit step (1.5). The OT combined with (1.5) is referred to as the optimal k-thresholding pursuit (OTP), and the ROT algorithm combined with (1.5) is called the relaxed optimal k-thresholding pursuit (ROTP), which are described, respectively, as follows.

OTP algorithm. Input (A, y, k). Given an initial point $x^0 \in \mathbb{R}^n$, repeat the following steps until a stoping criterion is satisfied:

S1. At x^p, set $u^p = x^p + A^T(y - Ax^p)$ and solve the binary optimization problem (3.5); let $w^*(u^p)$ be a solution of this problem.

S2. Set $S^{p+1} := \mathrm{supp}(u^p \otimes w^*(u^p))$, and let x^{p+1} be a solution to the problem

$$\min_x \{\|y - Ax\|_2^2 : \ \mathrm{supp}(x) \subseteq S^{p+1}\}.$$

ROTP algorithm. Input (A, y, k). Given an initial point $x^0 \in \mathbb{R}^n$, repeat the following steps until a stoping criterion is satisfied:

S1. At x^p, set $u^p = x^p + A^T(y - Ax^p)$, and solve the convex optimization problem

$$\min_w \{\|y - A(u^p \otimes w)\|_2^2 : \ \ \mathbf{e}^T w = k, \ 0 \le w \le \mathbf{e} \ \}$$

to generate a solution w^p of this problem.

S2. Let $v \in \mathcal{H}_k(u^p \otimes w^p)$, $S^{p+1} = \mathrm{supp}(v)$, and let x^{p+1} be a solution to the problem

$$\min_x \{\|y - Ax\|_2^2 : \ \mathrm{supp}(x) \subseteq S^{p+1}\}.$$

The vector w^p generated by the "compressing step" of the ROTP might not be sparse enough. This motivates the following enhanced versions of the ROTP called ROTP2 and ROTP3 which perform compressions of the data u^p two and three times, respectively, before the operator $\mathcal{H}_k$ is applied to the resulting compressible vector. As shown by numerical experiments (see section 5 for details), the aforementioned drawback of the hard thresholding will be remarkably overcome through compressing u^p more than once.

ROTP2 algorithm. Input (A, y, k). Given an initial point $x^0 \in R^n$, repeat the following steps until a stoping criterion is satisfied:

S1. At x^p, set $u^p = x^p + A^T(y - Ax^p)$ and solve the problem

$$\min_w \{\|y - A(u^p \otimes w)\|_2^2 : \ \ \mathbf{e}^T w = k, \ 0 \le w \le \mathbf{e} \ \}$$

to get a solution $w^{(1)}$ to this problem. Then solve the problem

$$\min_w\{\|y - A(u^p \otimes w^{(1)} \otimes w)\|_2^2 : \ \mathbf{e}^T w = k,\ 0 \le w \le \mathbf{e}\ \}$$

to get a solution $w^{(2)}$ to this problem.

S2. Let $v \in \mathcal{H}_k(u^p \otimes w^{(1)} \otimes w^{(2)})$ and $S^{p+1} = \mathrm{supp}(v)$. Let x^{p+1} be a solution to the problem

$$\min_x\{\|y - Ax\|_2^2 : \ \mathrm{supp}(x) \subseteq S^{p+1}\}.$$

ROTP3 algorithm. Input (A, y, k). Given an initial point $x^0 \in R^n$, repeat the following steps until a stoping criterion is satisfied:

S1. At x^p, set $u^p = x^p + A^T(y - Ax^p)$ and solve the problem

$$\min_w\{\|y - A(u^p \otimes w)\|_2^2 : \ \mathbf{e}^T w = k,\ 0 \le w \le \mathbf{e}\ \}$$

to get a solution $w^{(1)}$. Then solve

$$\min_w\{\|y - A(u^p \otimes w^{(1)} \otimes w)\|_2^2 : \ \mathbf{e}^T w = k,\ 0 \le w \le \mathbf{e}\ \}$$

to obtain a solution $w^{(2)}$, and then solve

$$\min_w\{\|y - A(u^p \otimes w^{(1)} \otimes w^{(2)} \otimes w)\|_2^2 : \ \mathbf{e}^T w = k,\ 0 \le w \le \mathbf{e}\ \}$$

to obtain a solution $w^{(3)}$.

S2. Let $v \in \mathcal{H}_k(u^p \otimes w^{(1)} \otimes w^{(2)} \otimes w^{(3)})$ and $S^{p+1} = \mathrm{supp}(v)$. Let x^{p+1} be the solution to the problem

$$\min_x\{\|y - Ax\|_2^2 : \ \mathrm{supp}(x) \subseteq S^{p+1}\}.$$

Before discussing numerical results, we prove the convergence of the basic algorithms presented in this section.

4. Theoretical performance. In this section, we establish the bound for the error of approximating the solution of (1.1) with the iterates generated by the OT, OTP, ROT, or ROTP under the RIP. In compressed sensing language, we prove the success of signal recovery via these algorithms under the RIP. Our analysis allows the measurements of the signal to be inaccurate, and we will point out at the end of this section that our analysis is also valid when the target signal x^* is not precisely k-sparse. In particular, if the measurements are accurate and the target signal is k-sparse, our results claim that the sequences generated by OT, OTP, ROT, or ROTP converge to the target signal under the RIP. Let us first recall the restricted isometry constant δ_K introduced by Candès and Tao [15].

DEFINITION 4.1 (see [15, 14]). *Given a matrix $A \in \mathbb{R}^{m\times n}$ with $m < n$, the Kth restricted isometry constant, denoted by δ_K, is the smallest number $\delta \ge 0$ such that*

$$(1 - \delta)\|x\|_2^2 \le \|Ax\|_2^2 \le (1 + \delta)\|x\|_2^2$$

holds for all K-sparse vector $x \in \mathbb{R}^n$.

The following properties will be frequently used in later analysis.

LEMMA 4.2 (see [15, 47, 30]). *Given $u \in \mathbb{R}^n$ and the set $S \subseteq \{1, 2, \ldots, n\}$, one has*

(i) $\|((I - A^T A)v)_S\|_2 \le \delta_t\|u\|_2$ *if* $|S \cup \mathrm{supp}(v)| \le t$.

(ii) $\|(A^T u)_S\|_2 \le \sqrt{1 + \delta_t}\|u\|_2$ *if* $|S| \le t$.

4.1. Analysis of OT and OTP algorithms. We first analyze the theoretical performance of the OT and OTP which provide a basic framework for the development of the ROT and ROTP and their variants. The main result for OT and OTP is summarized as follows.

THEOREM 4.3. *For every k-sparse vector x satisfying $y = Ax+\nu$, if the restricted isometry constant of the matrix A satisfies $\delta_{2k} < \tau^* \approx 0.5349$, where τ^* is the real root of the univariate equation $\tau^3 + \tau^2 + \tau = 1$, then the iterates $\{x^p\}$ generated by OT or OTP satisfy that*

$$\|x^p - x\|_2 \le \rho^p \|x^0 - x\|_2 + C\|\nu\|_2, \tag{4.1}$$

where ρ and C are constants given by

$$\rho = \delta_{2k}\sqrt{\frac{1+\delta_{2k}}{1-\delta_{2k}}} < 1, \quad C = \frac{3+\delta_{2k}}{(1-\rho)\sqrt{1-\delta_{2k}}}.$$

In particular, when $\nu = 0$, i.e., $y = Ax$, the iterates $\{x^p\}$ generated by OT or OTP converge to x.

Proof. Let x^p be the current iterate, generated by OT or OTP, which is k-sparse. Denote by $u^p = x^p + A^T(y - Ax^p)$ and $\mathcal{W}^{(k)} = \{w : \ \mathbf{e}^T w = k, \ w \in \{0,1\}^n\}$. Note that $y = Ax + \nu$. So

$$x - u^p = (I - A^T A)(x - x^p) - A^T\nu. \tag{4.2}$$

(I) We first analyze the OT algorithm. Note that $w^*(u^p) \in \mathcal{W}^{(k)}$ is a minimizer of the problem (3.5). Thus

$$\|y - A[u^p \otimes w^*(u^p)]\|_2 \le \|y - A(u^p \otimes w)\|_2 \ \text{ for any } w \in \mathcal{W}^{(k)}. \tag{4.3}$$

By the structure of the OT algorithm, $x^{p+1} = u^p \otimes w^*(u^p)$, and thus x^{p+1} is a k-sparse vector with $\mathrm{supp}(x^{p+1}) \subseteq \mathrm{supp}(w^*(u^p))$. Since x is a k-sparse vector, there exists a k-sparse binary vector $\widehat{w} \in \mathcal{W}^{(k)}$ such that $\mathrm{supp}(x) \subseteq \mathrm{supp}(\widehat{w})$, and hence $x \otimes (\mathbf{e} - \widehat{w}) = 0$. Then it follows from (4.3) that

$$\|y - Ax^{p+1}\|_2 \le \|y - A(u^p \otimes \widehat{w})\|_2. \tag{4.4}$$

Note that $x^{p+1} - x$ is a $(2k)$-sparse vector. By Lemma 4.2, we have $\|A(x - x^{p+1})\| \ge \sqrt{1-\delta_{2k}}\|x - x^{p+1}\|$. Thus

$$\|y - Ax^{p+1}\|_2 = \|A(x - x^{p+1}) + \nu\|_2 \ge \sqrt{1-\delta_{2k}}\|x - x^{p+1}\|_2 - \|\nu\|_2.$$

Merging this inequality with (4.4) leads to

$$\|x^{p+1} - x\|_2 \le \frac{1}{\sqrt{1-\delta_{2k}}}(\|y - A(u^p \otimes \widehat{w})\|_2 + \|\nu\|_2). \tag{4.5}$$

We now estimate the right-hand side of (4.5). By the choice of $\widehat{w}$ and noting that $\mathrm{supp}(x) \subseteq \mathrm{supp}(\widehat{w})$, we see that $|\mathrm{supp}(\widehat{w}) \cup \mathrm{supp}(x - x^p)| \le 2k$. Therefore, by (4.2) and Lemma 4.2, we have

$$\begin{aligned}
\|(x - u^p) \otimes \widehat{w}\|_2 &= \big\|[(I - A^T A)(x - x^p)] \otimes \widehat{w} - (A^T\nu) \otimes \widehat{w}\big\|_2 \\
&\le \big\|[(I - A^T A)(x - x^p)]_{\mathrm{supp}(\widehat{w})}\big\|_2 + \|(A^T\nu)_{\mathrm{supp}(\widehat{w})}\|_2 \\
&\le \delta_{2k}\|x - x^p\|_2 + \sqrt{1+\delta_k}\|\nu\|_2. \qquad (4.6)
\end{aligned}$$

As $(x-u^p)\otimes\widehat{w}$ is a k-sparse vector, we obtain

$$\begin{aligned}\|y-A(u^p\otimes\widehat{w})\|_2 &= \|\nu+A(x-u^p\otimes\widehat{w})\|_2\\ &= \|\nu+A[(x-u^p)\otimes\widehat{w}+x\otimes(\mathbf{e}-\widehat{w})]\|_2\\ &= \|\nu+A[(x-u^p)\otimes\widehat{w}]\|_2\\ &\le \|\nu\|_2+\sqrt{1+\delta_k}\|(x-u^p)\otimes\widehat{w}\|_2\\ &\le \delta_{2k}\sqrt{1+\delta_k}\|x-x^p\|_2+(2+\delta_k)\|\nu\|_2. \qquad (4.7)\end{aligned}$$

The third equality above follows from the fact $x\otimes(\mathbf{e}-\widehat{w})=0$. The first inequality above follows from Definition 4.1 with the fact $(x-u^p)\otimes\widehat{w}$ being k-sparse. The last inequality follows from (4.6). Note that $\delta_k\le\delta_{2k}$. Combining (4.5) and (4.7) yields

(4.8)
$$\|x^{p+1}-x\|_2\le\delta_{2k}\sqrt{\frac{1+\delta_k}{1-\delta_{2k}}}\|x-x^p\|_2+\frac{3+\delta_k}{\sqrt{1-\delta_{2k}}}\|\nu\|_2\le\rho\|x-x^p\|_2+\frac{3+\delta_{2k}}{\sqrt{1-\delta_{2k}}}\|\nu\|_2,$$

where $\rho:=\delta_{2k}\sqrt{\frac{1+\delta_{2k}}{1-\delta_{2k}}}<1$ which is ensured by $\delta_{2k}<\tau^*$, where τ^* (≈ 0.5349) is the positive real root of the univariate equation $\tau^3+\tau^2+\tau=1$. The result (4.1) follows immediately from (4.8) and the fact $\sum_{i=1}^{\infty}\rho^i=\frac{1}{1-\rho}$.

(II) We now consider the OTP algorithm, which generates the next iterate x^{p+1} by performing the orthogonal project step

$$\min_z\{\|y-Az\|_2^2:\ \mathrm{supp}(z)\subseteq\mathrm{supp}(u^p\otimes w^*(u^p))\},$$

which implies that

$$\|y-Ax^{p+1}\|_2\le\|y-A(u^p\otimes w^*(u^p))\|_2\le\|y-A(u^p\otimes\hat{w})\|_2,$$

where the last inequality follows from (4.3) by setting $w=\widehat{w}$. Therefore, the iterate x^{p+1} generated by the OTP also satisfies the relation (4.4). Repeating the same proof above for the OT algorithm, we see that (4.8) remains valid for the OTP with the same constant $\rho<1$.

In particular, when $\nu=0$, it follows immediately from (4.1) that the sequence $\{x^p\}$ generated by OT or OTP converges to x. ∎

The result above is shown under the condition $\delta_{2k}<\tau^*$. The RIP condition has been widely used in the theoretical analysis of various thresholding algorithms. For instance, the convergence of the HTP was shown under the condition $\delta_{3k}<1/\sqrt{3}$ (see [30, 32]), and that of the IHT algorithm with a stepsize taken in $(\frac{1}{2(1-\delta_{2k})},\frac{1}{1+\delta_{2k}})$ was shown under the condition $\delta_{2k}<1/3$ (see [33, 9, 6, 32]).

In the case $\nu=0$, the convergence rate of $\{x^p\}$ in Theorem 4.3 can be further enhanced, as shown by the next corollary.

COROLLARY 4.4 (local convergence rate). *For every k-sparse vector x with $y=Ax$, under the same condition of Theorem 4.3, there exists an integer number $\widehat{p}$ such that for all $p\ge\widehat{p}$,*

$$\|x^p-x\|_2\le(\rho^*)^p\|x^0-x\|_2,$$

where

$$\rho^*:=\delta_k\sqrt{\frac{1+\delta_k}{1-\delta_k}}\le\rho=\delta_{2k}\sqrt{\frac{1+\delta_{2k}}{1-\delta_{2k}}}<1. \qquad (4.9)$$

Proof. By Theorem 4.3, when $\nu = 0$, the sequence $\{x^p\}$ generated by OT or OTP algorithm converges to x. Thus there is a sufficiently large integer number $\widehat{p}$ such that $\mathrm{supp}(x) \subseteq \mathrm{supp}(x^p)$ for any $p \geq \widehat{p}$. In fact, if there is an index $i_0 \in \mathrm{supp}(x)$ such that $i_0 \notin \mathrm{supp}(x^p)$, then $\|x - x^p\|_2 \geq |x_{i_0}|$, contradicting the fact $x^p \to x$ as $p \to \infty$. Therefore, $x - x^{p+1}$ and $x - x^p$ must be k-sparse for all $p \geq \hat{p}$. The left-hand side of (4.3) is larger than or equal to $\sqrt{1-\delta_k}\|x - x^{p+1}\|_2$. For $p \geq \widehat{p}$, pick a vector in $\mathcal{W}^{(k)}$, denoted by $\widehat{w}^p$, which satisfies that $\mathrm{supp}(x^p) \subseteq \mathrm{supp}(\widehat{w}^p)$. This implies that $\mathrm{supp}(x) \subseteq \mathrm{supp}(\widehat{w}^p)$ for all $p \geq \widehat{p}$. Therefore, $x \otimes (\mathbf{e} - \widehat{w}^p) = 0$ for all $p \geq \widehat{p}$. Replacing the vector $\widehat{w}$ in the proof of Theorem 4.3 with $\widehat{w}^p$, the inequality (4.6) can be improved to $\|(x - u^p) \otimes \widehat{w}^p\|_2 \leq \delta_k \|x - x^p\|_2$ due to the fact $\nu = 0$ and $|\mathrm{supp}(\widehat{w}^p) \cup \mathrm{supp}(x - x^p)| \leq k$. The estimation (4.7) can be improved to

$$\|y - A(u^p \otimes \widehat{w}^p)\|_2 \leq \delta_k \sqrt{1+\delta_k}\|x - x^p\|_2.$$

Therefore, from the proof of Theorem 4.3, we have

$$\|x^{p+1} - x\|_2 \leq \delta_k \sqrt{\frac{1+\delta_k}{1-\delta_k}}\|x - x^p\|_2, \ p \geq \widehat{p}.$$

Since $\delta_k \leq \delta_{2k} < \tau^*$, we immediately see the relation in (4.9). ∎

This result indicates that the local convergence speed of the OT and OTP may actually be faster than what Theorem 4.3 claims.

4.2. Analysis of ROT and ROTP algorithms. We now analyze the ROT and ROTP algorithms which are the tightest convex relaxation counterparts of the OT and OTP, respectively. Note that the solution w^p of the relaxation problem in (3.6) may not be exactly binary (and hence may not be k-sparse). So the analysis in section 4.1, based on the optimal k-thresholding indicator $w^*(u^p)$, does not apply to the ROT and ROTP for which a nontrivial analysis will be provided in this section. We first give a few useful lemmas.

LEMMA 4.5. *Let $z \in \mathbb{R}^n$ be a given vector. Then for any $\hat{z} \in \mathcal{H}_k(z)$, one has*

$$\|z - \hat{z}\|_2^2 \leq \|z - x\|_2^2 - \|(z - x)_S\|_2^2$$

for any k-sparse vector $x \in \mathbb{R}^n$ with $S = \mathrm{supp}(x)$.

Proof. Since $\mathcal{H}_k(z)$ retains the largest k magnitudes of z, for any $\hat{z} \in \mathcal{H}_k(z)$, $\|z - \hat{z}\|_2^2$ is the sum of the squares of the $n-k$ smallest magnitudes of z, which must be smaller than or equal to the sum of the squares of any $n-k$ components of z. So $\|z - \hat{z}\|_2^2 \leq \|z_{\overline{S}}\|_2^2$ for any set $S \subseteq \{1, \ldots, n\}$ with $|S| \leq k$, where $\overline{S} = \{1, \ldots, n\} \backslash S$. Let x be any k-sparse vector with $S = \mathrm{supp}(x)$. As $x_{\overline{S}} = 0$ and $|S| \leq k$, by setting $S = \mathrm{supp}(x)$ in the inequality above, we immediately have

$$\|z - \hat{z}\|_2^2 \leq \|z_{\overline{S}}\|_2^2 = \|(z - x)_{\overline{S}}\|_2^2 = \|z - x\|_2^2 - \|(z - x)_S\|_2^2,$$

as desired. ∎

LEMMA 4.6. *Let $u^p \in \mathbb{R}^n$ be a given vector, and let $x \in \mathbb{R}^n$ be a k-sparse vector with $S = \mathrm{supp}(x)$. Let w^p be a solution to the problem* (3.6)*. Then for any vector $v \in \mathcal{H}_k(u^p \otimes w^p)$ with $S^{p+1} = \mathrm{supp}(v)$, one has*

$$\|x - v\|_2 \leq \|(u^p \otimes w^p - x)_{S^{p+1} \cup S}\|_2 + \|(u^p \otimes w^p - x)_{S^{p+1} \backslash S}\|_2.$$

Proof. Let $x \in \mathbb{R}^n$ be a k-sparse vector with $S = \mathrm{supp}(x)$. By setting $z = u^p \otimes w^p$ in Lemma 4.5, for any $v \in \mathcal{H}_k(u^p \otimes w^p)$, we have

$$\|u^p \otimes w^p - v\|_2^2 \le \|u^p \otimes w^p - x\|_2^2 - \|(u^p \otimes w^p - x)_S\|_2^2.$$

The left-hand side can be written as

$$\|u^p \otimes w^p - v\|_2^2 = \|u^p \otimes w^p - x\|_2^2 + \|x - v\|_2^2 + 2(x-v)^T(u^p \otimes w^p - x).$$

Note that $\mathrm{supp}\,(x-v) \subseteq \mathrm{supp}(v) \cup \mathrm{supp}(x) = S^{p+1} \cup S$, where $S^{p+1} = \mathrm{supp}(v)$. Combining the two relations above yields

$$\begin{aligned} \|x - v\|_2^2 &\le -\|(u^p \otimes w^p - x)_S\|_2^2 - 2(x-v)^T(u^p \otimes w^p - x). \\ &= -\|(u^p \otimes w^p - x)_S\|_2^2 - 2[(x-v)_{S^{p+1}\cup S}]^T[u^p \otimes w^p - x]_{S^{p+1}\cup S} \\ &\le -\|(u^p \otimes w^p - x)_S\|_2^2 + 2\|x-v\|_2\|[u^p \otimes w^p - x]_{S^{p+1}\cup S}\|_2. \end{aligned} \tag{4.10}$$

Note that the positive root of the quadratic function (in variable t)

$$t^2 - 2t\|(u^p \otimes w^p - x)_{S^{p+1}\cup S}\|_2 + \|(u^p \otimes w^p - x)_S\|_2^2 = 0$$

is given as follows:

$$\begin{aligned} t^* &= \frac{2\|(u^p \otimes w^p - x)_{S^{p+1}\cup S}\|_2 + \sqrt{4\|(u^p \otimes w^p - x)_{S^{p+1}\cup S}\|_2^2 - 4\|(u^p \otimes w^p - x)_S\|_2^2}}{2} \\ &= \|(u^p \otimes w^p - x)_{S^{p+1}\cup S}\|_2 + \|(u^p \otimes w^p - x)_{S^{p+1}\setminus S}\|_2 \end{aligned}$$

The inequality (4.10) implies that $\|x - v\|_2 \le t^*$, as desired. ∎

The next lemma has been shown in the proof of Theorem 4.3. See (4.7) for details.

LEMMA 4.7. *Let $x \in \mathbb{R}^n$ be a k-sparse vector satisfying $y = Ax + \nu$. Let $x^p \in \mathbb{R}^n$ and $u^p = x^p + A^T(y - Ax^p)$. Then for any $\widehat{w} \in \mathcal{W}^{(k)}$ satisfying $supp(x) \subseteq \widehat{w}$, one has*

$$\|y - A(u^p \otimes \widehat{w})\|_2 \le \delta_{2k}\sqrt{1+\delta_k}\|x - x^p\|_2 + (2+\delta_k)\|\nu\|_2.$$

We now prove the main result for ROT and ROTP algorithms.

THEOREM 4.8. *Let x be a k-sparse vector satisfying $y = Ax + \nu$. Suppose that the restricted isometry constant of the matrix A satisfies $\delta_{3k} \le 1/5$. Then the iterates $\{x^p\}$, generated by ROT or ROTP, approximate x with error*

$$\|x^p - x\|_2 \le \varrho^p\|x^0 - x\|_2 + C^*\|\nu\|_2, \tag{4.11}$$

where, for ROT, the constants ϱ and C^* are given as

$$\varrho := (\delta_{2k} + 2\delta_{3k})\sqrt{\frac{1+\delta_k}{1-\delta_{2k}}} + \delta_{3k} < 1, \quad C^* = \frac{1}{1-\varrho}\left(\frac{5+3\delta_k}{\sqrt{1-\delta_{2k}}} + \sqrt{1+\delta_k}\right),$$

and for ROTP the constants ϱ and C^* are given as

$$\varrho = \frac{1}{\sqrt{1-\delta_{2k}^2}}\left((\delta_{2k} + 2\delta_{3k})\sqrt{\frac{1+\delta_k}{1-\delta_{2k}}} + \delta_{3k}\right) < 1, \tag{4.12}$$

$$C^* = \frac{1}{1-\varrho}\left(\frac{5+3\delta_k}{(1-\delta_{2k})\sqrt{1+\delta_{2k}}} + \frac{\sqrt{1+\delta_k}}{\sqrt{1-\delta_{2k}^2}} + \frac{\sqrt{1+\delta_k}}{1-\delta_{2k}}\right). \tag{4.13}$$

In particular, when $\nu = 0$ (i.e., $y = Ax$), the sequence $\{x^p\}$ generated by ROT or ROTP converges to x.

Proof. (I) We first analyze the ROT. At x^p, the ROT generates the vector w^p by solving the optimization problem (3.6) with $u^p = x^p + A^T(y - Ax^p)$. Then the next iterate is given by $x^{p+1} \in \mathcal{H}_k(u^p \otimes w^p)$. Denote by $S^{p+1} = \mathrm{supp}(x^{p+1})$. Since x is a k-sparse vector with $S = \mathrm{supp}(x)$, by Lemma 4.6, we have

$$\|x - x^{p+1}\|_2 \le \|(u^p \otimes w^p - x)_{S^{p+1}\cup S}\|_2 + \left\|(u^p \otimes w^p - x)_{S^{p+1}\setminus S}\right\|_2. \tag{4.14}$$

We now estimate the upper bound for the right-hand side of the above inequality. By using (4.2) and noting that $x_{S^{p+1}\setminus S} = 0$ and $0 \le w^p \le \mathbf{e}$, we have

$$\begin{aligned}\left\|(u^p \otimes w^p - x)_{S^{p+1}\setminus S}\right\|_2 &= \left\|(u^p \otimes w^p)_{S^{p+1}\setminus S}\right\|_2 = \|[(u^p - x) \otimes w^p]_{S^{p+1}\setminus S}\|_2 \\ &= \left\|[(A^T\nu - (I - A^TA)(x - x^p)) \otimes w^p]_{S^{p+1}\setminus S}\right\|_2 \\ &\le \left\|[(I - A^TA)(x - x^p)]_{S^{p+1}\setminus S}\right\|_2 + \|(A^T\nu)_{S^{p+1}\setminus S}\|_2 \\ &\le \delta_{3k}\|x^p - x\|_2 + \sqrt{1+\delta_k}\|\nu\|_2, \end{aligned} \tag{4.15}$$

where the last inequality follows from Lemma 4.2 due to the fact $|\mathrm{supp}(x - x^p) \cup (S^{p+1}\setminus S)| \le 3k$ and $|S^{p+1}\setminus S| \le k$. Using $y = Ax + \nu$, we have

$$\begin{aligned}&\|y - A(u^p \otimes w^p)\|_2 \\ &= \|A(u^p \otimes w^p - x) - \nu\|_2 \\ &= \left\|A[(u^p \otimes w^p - x)_{S^{p+1}\cup S}] + A[(u^p \otimes w^p - x)_{\overline{S^{p+1}\cup S}}] - \nu\right\|_2 \\ &\ge \|A[(u^p \otimes w^p - x)_{S^{p+1}\cup S}]\|_2 - \left\|A[(u^p \otimes w^p - x)_{\overline{S^{p+1}\cup S}}]\right\|_2 - \|\nu\|_2 \\ &\ge \sqrt{1-\delta_{2k}}\|(u^p \otimes w^p - x)_{S^{p+1}\cup S}\|_2 - \left\|A[(u^p \otimes w^p - x)_{\overline{S^{p+1}\cup S}}]\right\|_2 - \|\nu\|_2, \end{aligned}$$

and thus

$$\|(u^p \otimes w^p - x)_{S^{p+1}\cup S}\|_2 \le \frac{1}{\sqrt{1-\delta_{2k}}}\left(\|y - A(u^p \otimes w^p)\|_2 + \mathcal{T} + \|\nu\|_2\right), \tag{4.16}$$

where

$$\mathcal{T} := \|A[(u^p \otimes w^p - x)_{\overline{S^{p+1}\cup S}}]\|_2.$$

Let $\widehat{w} \in \mathcal{W}^{(k)}$ be a vector such that $S = \mathrm{supp}(x) \subseteq \mathrm{supp}(\widehat{w})$, which implies that $x \otimes (\mathbf{e} - \widehat{w}) = 0$. Since w^p is an optimal solution to (3.6), we have

$$\|y - A(u^p \otimes w^p)\|_2 \le \|y - A(u^p \otimes \widehat{w})\|_2 \le \delta_{2k}\sqrt{1+\delta_k}\|x^p - x\|_2 + (2+\delta_k)\|\nu\|_2, \tag{4.17}$$

where the last inequality follows from Lemma 4.7. Combining (4.14), (4.15), (4.16), and (4.17), we have

$$\begin{aligned}\|x - x^{p+1}\|_2 \le{}& \frac{1}{\sqrt{1-\delta_{2k}}}\left[\delta_{2k}\sqrt{1+\delta_k}\|x^p - x\|_2 + \mathcal{T}\right] + \left[\frac{3+\delta_k}{\sqrt{1-\delta_{2k}}} + \sqrt{1+\delta_k}\right]\|\nu\|_2 \\ &+\delta_{3k}\|x^p - x\|_2. \end{aligned} \tag{4.18}$$

In the remainder of the proof, we estimate the term $\mathcal{T}$. Since $x_{\overline{S\cup S^{p+1}}} = 0$, $\mathcal{T}$ can be written as

$$\mathcal{T} = \|A[(u^p \otimes w^p)_{\overline{S^{p+1}\cup S}}]\|_2 = \|A[(u^p - x) \otimes w^p]_{\overline{S^{p+1}\cup S}}\|_2.$$

Let $|\overline{S^{p+1} \cup S}| = (\widehat{n} - 1)k + \ell$, where $\widehat{n}$ and ℓ are integer numbers and $0 \le \ell < k$. Let

$$\overline{S^{p+1} \cup S} = S_1 \cup S_2 \cup \cdots \cup S_{\widehat{n}-1} \cup S_{\widehat{n}}$$

be the disjoined partition of $\overline{S^{p+1} \cup S}$, satisfying the following properties:

(i) $S_i \cap S_j = \emptyset$ for $i \neq j$, and $|S_i| = k$ for all $i = 1, \ldots, \widehat{n} - 1$ and $|S_{\widehat{n}}| = \ell < k$.

(ii) S_1 is the index set for the k largest elements in the set $\{(w^p)_i : i \in \overline{S^{p+1} \cup S}\}$, S_2 is the index set for the second k largest elements in this set, and so on.

Thus the vector $(w^p)_{\overline{S^{p+1} \cup S}}$ is decomposed as

$$(w^p)_{\overline{S^{p+1} \cup S}} = (w^p)_{S_1} + \cdots + (w^p)_{S_{\widehat{n}-1}} + (w^p)_{S_{\widehat{n}}}.$$

Sort the components of w^p supported on S_i $(i = 1, \ldots, \hat{n} - 1)$ into descending order, and denote such ordered components by $\alpha_1^{(i)} \geq \alpha_2^{(i)} \geq \cdots \geq \alpha_k^{(i)}$, and denote the ordered components of w^p supported on $S_{\hat{n}}$ by $\alpha_1^{(\hat{n})} \geq \alpha_2^{(\hat{n})} \geq \cdots \geq \alpha_\ell^{(\hat{n})}$. Thus $\alpha_1^{(i)}$ denotes the largest entries of w^p on the support S_i for $i = 1, \ldots, \hat{n}$, $\alpha_k^{(i)}$ denotes the smallest entry of w^p on the support S_i for $i = 1, \ldots, \hat{n} - 1$, and $\alpha_\ell^{(\hat{n})}$ denotes the smallest component of w^p supported on $S_{\hat{n}}$. By this notation, sorting the components of the vector $(w^p)_{\overline{S^{p+1} \cup S}}$ supported on $\overline{S^{p+1} \cup S}$ into descending order, we obtain the sequence as follows:

$$\overbrace{\alpha_1^{(1)} \geq \alpha_2^{(1)} \geq \cdots \geq \alpha_k^{(1)}} \geq \overbrace{\alpha_1^{(2)} \geq \alpha_2^{(2)} \geq \cdots \geq \alpha_k^{(2)}} \geq \cdots \geq \overbrace{\alpha_1^{(\widehat{n})} \geq \alpha_2^{(\widehat{n})} \geq \cdots \geq \alpha_\ell^{(\widehat{n})}}.$$

We now prove that

$$\Delta := \sum_{i=1}^{\widehat{n}} \alpha_1^{(i)} \leq 2 - \frac{1}{k} < 2. \tag{4.19}$$

For each i, the largest entry of $(w^p)_{S_{i+1}}$ is smaller than or equal to the smallest entry of $(w^p)_{S_i}$, i.e., $\alpha_k^{(i)} \geq \alpha_1^{(i+1)}$ for $i = 1, \ldots, \widehat{n} - 1$. So we immediately see that

$$\Delta = \alpha_1^{(1)} + \alpha_1^{(2)} + \cdots + \alpha_1^{(\widehat{n})} \leq \alpha_1^{(1)} + \alpha_k^{(1)} + \cdots + \alpha_k^{(\widehat{n}-1)} \leq 1 + \sum_{i=1}^{\widehat{n}-1} \alpha_k^{(i)},$$

where the last inequality follows from $\alpha_1^{(1)} \leq 1$ due to the fact $0 \leq w^p \leq \mathbf{e}$. Note that $\alpha_k^{(i)} \leq \alpha_{k-1}^{(i)} \leq \cdots \leq \alpha_2^{(i)}$. Thus

$$\sum_{i=1}^{\widehat{n}-1} \alpha_k^{(i)} \leq \sum_{i=1}^{\widehat{n}-1} \alpha_{k-1}^{(i)} \leq \cdots \leq \sum_{i=1}^{\widehat{n}-1} \alpha_2^{(i)}.$$

So it follows from the inequalities above that $\Delta \leq 1 + \sum_{i=1}^{\widehat{n}-1} \alpha_j^{(i)}$ for $j = 2, \ldots, k$. Adding these $k - 1$ inequalities to the equality $\Delta = \sum_{i=1}^{\widehat{n}} \alpha_1^{(i)}$ yields

$$\begin{aligned} k\Delta &\leq k - 1 + \sum_{i=1}^{\widehat{n}} \alpha_1^{(i)} + \sum_{i=1}^{\widehat{n}-1} \alpha_2^{(i)} + \cdots + \sum_{i=1}^{\widehat{n}-1} \alpha_k^{(i)} \leq k - 1 + \sum_{j \in \overline{S^{p+1} \cup S}} (w^p)_j \\ &\leq k - 1 + \|w^p\|_1 \\ &= 2k - 1, \end{aligned}$$

where the last inequality follows from the fact $\|w^p\|_1 = \mathbf{e}^T w^p = k$. Thus (4.19) holds. Define the vector $v^{(i)} := [(u^p - x) \otimes w^p]_{S_i}$; then

$$[(u^p - x) \otimes w^p]_{\overline{S^{p+1} \cup S}} = v^{(1)} + v^{(2)} + \cdots + v^{(\widehat{n})}.$$

So the vector $[(u^p - x) \otimes w^p]_{\overline{S^{p+1} \cup S}}$ is decomposed into k-sparse vectors $v^{(i)} \in \mathbb{R}^n, i = 1, \ldots, \widehat{n}$. Therefore,

$$\mathcal{T} = \left\| A \sum_{i=1}^{\widehat{n}} v^{(i)} \right\|_2 \le \sum_{i=1}^{\widehat{n}} \|Av^{(i)}\|_2 \le \sqrt{1+\delta_k} \sum_{i=1}^{\widehat{n}} \|v^{(i)}\|_2, \tag{4.20}$$

where the last inequality follows from Definition 4.1 and the fact that every $v^{(i)}$ is k-sparse. We now estimate the term $\sum_{i=1}^{\widehat{n}} \|v^{(i)}\|_2$. Note that

$$\begin{aligned}
\|v^{(i)}\|_2 &= \|[(u^p - x) \otimes w^p]_{S_i}\|_2 \\
&= \|[(A^T\nu) \otimes w^p - ((I - A^TA)(x - x^p)) \otimes w^p]_{S_i}\|_2 \\
&\le \|[(I - A^TA)(x - x^p)]_{S_i} \otimes (w^p)_{S_i}\|_2 + \|(A^T\nu)_{S_i} \otimes (w^p)_{S_i}\|_2 \\
&\le \left(\max_{i \in S_i} (w^p)_i\right) \|[(I - A^TA)(x - x^p)]_{S_i}\|_2 + \left(\max_{i \in S_i} (w^p)_i\right) \|(A^T\nu)_{S_i}\|_2 \\
&\le \alpha_1^{(i)} \delta_{3k} \|x - x^p\|_2 + \alpha_1^{(i)} \sqrt{1+\delta_k} \|\nu\|_2,
\end{aligned}$$

where the last inequality follows from the fact $\alpha_1^{(i)}$ being the largest entry of $(w^p)_{S_i}$ and Lemma 4.2 with $|S_i \cup \mathrm{supp}(x - x^p)| \le 3k$. Thus

$$\sum_{i=1}^{\widehat{n}} \|v^{(i)}\|_2 \le \Delta\delta_{3k} \|x - x^p\|_2 + \Delta\sqrt{1+\delta_k} \|\nu\|_2 \le 2\delta_{3k} \|x - x^p\|_2 + 2\sqrt{1+\delta_k} \|\nu\|_2.$$

Merging (4.20) and the inequality above leads to

$$\mathcal{T} \le 2\delta_{3k}\sqrt{1+\delta_k} \|x - x^p\|_2 + 2(1+\delta_k) \|\nu\|_2.$$

Combining (4.18) and the above bound of $\mathcal{T}$ yields

$$\|x - x^{p+1}\|_2 \le \varrho \|x - x^p\|_2 + \left[\frac{5 + 3\delta_k}{\sqrt{1-\delta_{2k}}} + \sqrt{1+\delta_k}\right] \|\nu\|_2, \tag{4.21}$$

where

$$\varrho := (\delta_{2k} + 2\delta_{3k}) \sqrt{\frac{1+\delta_k}{1-\delta_{2k}}} + \delta_{3k} < 1 \tag{4.22}$$

under the condition $\delta_{3k} \le 1/5$. In fact, since $\delta_k \le \delta_{2k} \le \delta_{3k}$, we see that $\varrho \le 3\delta_{3k}\sqrt{\frac{1+\delta_{3k}}{1-\delta_{3k}}} + \delta_{3k} < 1$ which is ensured by the condition $\delta_{3k} \le 1/5$. The bound (4.11) immediately follows from (4.21) and (4.22).

(II) We now analyze the ROTP algorithm under the same assumption. The ROTP solves the same optimization problem (3.6) to obtain the vector w^p. Let v be an arbitrary vector in $\mathcal{H}_k(u^p \otimes w^p)$. In ROT, v is directly taken as the next iterate x^{p+1}. The bound (4.21), which is shown for ROT, holds for any vector v in $\mathcal{H}_k(u^p \otimes w^p)$. Therefore,

$$\|x - v\|_2 \le \varrho \|x - x^p\|_2 + C' \|\nu\|_2, \tag{4.23}$$

where ϱ is given by (4.22) and $C' = \frac{5+3\delta_k}{\sqrt{1-\delta_{2k}}} + \sqrt{1+\delta_k}$. The ROTP uses v as the intermediate point to compute the iterate x^{k+1} which is the solution to the orthogonal projection problem

$$\min_z \{\|y - Az\|_2^2 : \ \mathrm{supp}(z) \subseteq S^{p+1} = \mathrm{supp}(v)\}.$$

Thus by optimality, the solution x^{p+1} to this problem must satisfy that $[A^T(y - Ax^{p+1})]_{S^{p+1}} = 0$ which, by using $y = Ax + \nu$, can be written as

$$[(I - A^TA)(x - x^{p+1})]_{S^{p+1}} = (x - x^{p+1})_{S^{p+1}} + (A^T\nu)_{S^{p+1}}.$$

This implies that

$$\begin{aligned}\|(x - x^{p+1})_{S^{p+1}}\|_2 &\le \|[(I - A^TA)(x - x^{p+1})]_{S^{p+1}}\|_2 + \|(A^T\nu)_{S^{p+1}}\|_2 \\ &\le \delta_{2k}\|x - x^{p+1}\|_2 + \sqrt{1+\delta_k}\|\nu\|_2.\end{aligned}$$

The last equality follows from Lemma 4.2 due to the fact $|\mathrm{supp}(x-x^{p+1})\cup S^{p+1}| \le 2k$ and $|S^{p+1}| \le k$. Noting that $(x^{p+1})_{\overline{S^{p+1}}} = 0$ and $v_{\overline{S^{p+1}}} = 0$, we have

$$\begin{aligned}\|x - x^{p+1}\|_2^2 &= \|(x - x^{p+1})_{S^{p+1}}\|_2^2 + \|(x - x^{p+1})_{\overline{S^{p+1}}}\|_2^2 \\ &= \|(x - x^{p+1})_{S^{p+1}}\|_2^2 + \|(x - v)_{\overline{S^{p+1}}}\|_2^2 \\ &\le \delta_{2k}^2\|x - x^{p+1}\|_2^2 + 2\delta_{2k}\sqrt{1+\delta_k}\|x - x^{p+1}\|_2\|\nu\|_2 + (1+\delta_k)\|\nu\|_2^2 \\ &\quad + \|(x - v)_{\overline{S^{p+1}}}\|_2^2,\end{aligned}$$

and hence

$$(1-\delta_{2k}^2)\|x - x^{p+1}\|_2^2 \le 2\delta_{2k}\sqrt{1+\delta_k}\|x - x^{p+1}\|_2\|\nu\|_2 + (1+\delta_k)\|\nu\|_2^2 + \|(x - v)_{\overline{S^{p+1}}}\|_2^2.$$

This implies that

$$\begin{aligned}\|x - x^{p+1}\|_2 &\le \frac{2\delta_{2k}\sqrt{1+\delta_k}\|\nu\|_2 + \sqrt{4(1+\delta_k)\|\nu\|_2^2 + 4(1-\delta_{2k}^2)\|(x - v)_{\overline{S^{p+1}}}\|_2^2}}{2(1-\delta_{2k}^2)} \\ &\le \frac{2\delta_{2k}\sqrt{1+\delta_k}\|\nu\|_2 + 2\sqrt{1+\delta_k}\|\nu\|_2 + 2\sqrt{1-\delta_{2k}^2}\|(x - v)_{\overline{S^{p+1}}}\|_2}{2(1-\delta_{2k}^2)} \\ &\le \frac{\sqrt{1+\delta_k}}{1-\delta_{2k}}\|\nu\|_2 + \frac{1}{\sqrt{1-\delta_{2k}^2}}\|(x - v)_{\overline{S^{p+1}}}\|_2 \\ &\le \frac{\sqrt{1+\delta_k}}{1-\delta_{2k}}\|\nu\|_2 + \frac{1}{\sqrt{1-\delta_{2k}^2}}\|x - v\|_2.\end{aligned}$$

Combining this inequality with (4.23) yields

$$\|x - x^{p+1}\|_2 \le \frac{\varrho\|x - x^p\|_2}{\sqrt{1-\delta_{2k}^2}} + \left[\frac{C'}{\sqrt{1-\delta_{2k}^2}} + \frac{\sqrt{1+\delta_k}}{1-\delta_{2k}}\right]\|\nu\|_2 = \varrho'\|x - x^p\|_2 + C''\|\nu\|_2,$$

where

$$C'' = \frac{5 + 3\delta_k}{(1-\delta_{2k})\sqrt{1+\delta_{2k}}} + \frac{\sqrt{1+\delta_k}}{\sqrt{1-\delta_{2k}^2}} + \frac{\sqrt{1+\delta_k}}{1-\delta_{2k}}$$

and

$$\varrho' := \frac{1}{\sqrt{1-\delta_{2k}^2}}\left[(\delta_{2k} + 2\delta_{3k})\sqrt{\frac{1+\delta_k}{1-\delta_{2k}}} + \delta_{3k}\right] \le \frac{3\delta_{3k}}{1-\delta_{3k}} + \frac{\delta_{3k}}{\sqrt{1-\delta_{3k}^2}} < 1,$$

where the first inequality follows from the fact $\delta_k \le \delta_{2k} \le \delta_{3k}$, and the last one follows from the condition $\delta_{3k} \le 1/5$. Thus the error bound (4.11), with constants (4.12) and (4.13), holds for ROTP.

In particular, when $\nu = 0$, i.e., $y = Ax$, the iterates $\{x^p\}$ generated by the ROT and ROTP converge to the sparse vector x. ∎

Remark. In signal recovery scenarios, the target signal x is usually not exactly k-sparse and the measurements $y = Ax + \phi$ are also inaccurate, where ϕ is a noise vector. In such situations, we are interested in recovering the k largest magnitudes of x (which usually carry the most important information of the signal). Our main results (Theorems 4.3 and 4.8) can be immediately applied to such situations. In fact, let $S \subseteq \{1, \ldots, n\}$ denote the index set for the k largest magnitudes of the target signal x. Note that

$$y = Ax + \phi = Ax_S + (Ax_{\overline{S}} + \phi) = Ax_S + \nu,$$

where $\nu = Ax_{\overline{S}} + \phi$ and $\overline{S} = \{1, \ldots, n\} \backslash S$. The measurements y of the original signal x with noise ϕ can be seen as the measurements of the k-sparse vector x_S with noise $\nu = Ax_{\overline{S}} + \phi$. Therefore, Theorem 4.3 claims that if $\delta_{2k} < \tau^* \approx 0.5349$, then the iterates $\{x^p\}$ generated by OT or OTP approximate x_S with error

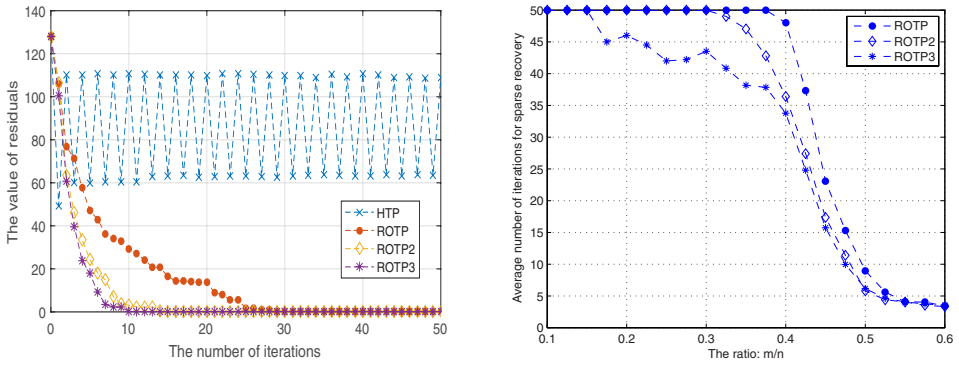
$$\|x^p - x_S\|_2 \leq \rho^p \|x^0 - x_S\|_2 + C\|Ax_{\overline{S}} + \phi\|_2, \tag{4.24}$$

where ρ and C are constants given in Theorem 4.3. Also Theorem 4.8 shows that if $\delta_{3k} \leq 1/5$, then the iterate x^p generated by ROT or ROTP approximates x_S with the error (4.24), where the constants ρ and C are replaced, respectively, with ϱ and C^* that are given in Theorem 4.8.

5. Numerical performance. Some preliminary experiments were performed to demonstrate the numerical behavior of the proposed algorithms. All matrices and sparse vectors are randomly generated. The entries of matrices are assumed to be independently and identically distributed random variables which follow $\mathcal{N}(0, 1)$, the standard normal distribution with zero mean and unit variance. The nonzero entries of the sparse vectors realized in our experiments are also assumed to follow such a distribution, and the random positions of nonzero entries are chosen according to a uniform distribution. All experiments were performed on a PC with the processor Intel(R) Core(TM) i5-3570 CPU @ 3.40 GHz and 8GB memory. All programs were written in MATLAB, and the convex optimization problems were solved by using CVX developed by Grant and Boyd [34] with solver "sedumi."

The first experiment was performed to illustrate the stableness of the proposed algorithms with respect to residual reduction. We generate a random matrix $A \in \mathbb{R}^{500\times 1000}$ and a random sparse vector $x^* \in \mathbb{R}^{1000}$ with sparsity level $k = 120$ (i.e., $\|x\|_0 \leq 120$) and then set $y := Ax^*$. We perform the HTP, ROTP, ROTP2, and ROTP3 up to 50 iterations, and the values of the residual $\|y - Ax^p\|_2$ with respect to the number of iterations for these algorithms are described in Figure 5.1(a). It can clearly be seen that our algorithms are stable in the sense that the residual is successively reduced to the prescribed tolerance $\|y - Ax^p\|_2 \leq 10^{-8}$ within a small number of iterations. From Figure 5.1(a), however, the residuals at the iterates generated by the HTP oscillate dramatically with no clear movement towards the solution of the problem over the course of iterations. This oscillation phenomenon in HTPs was not observed in the ROTP and its enhanced versions, although such experiments were repeated a number of times on random examples of the problems. This experiment also indicates that the number of iterations required by the ROTP2 and ROTP3 to find the solution of a problem is lower than the number of iterations required by the ROTP. This means compressing the vector $u^p = x^p + A^T(y - Ax^p)$ more than once does improve the stability and efficiency of the algorithm, as predicted in section 3.

The second experiment was performed to demonstrate the average number of iterations required by the proposed algorithms to meet a prescribed recovery criterion.

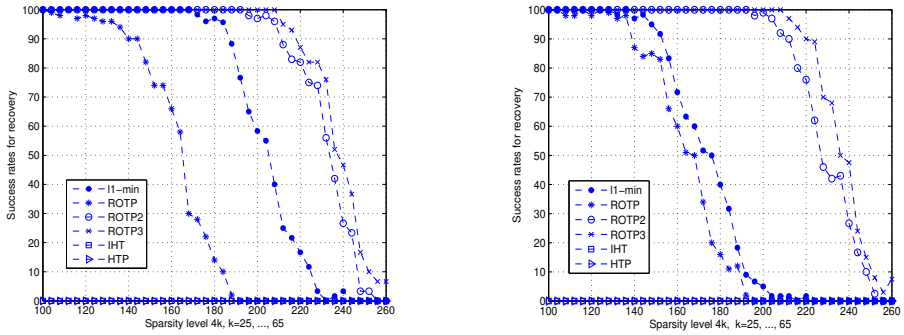


(a) Residual reduction (b) Average no. of iter. for recovery

FIG. 5.1. *Comparision of several algorithms in residual reduction and the average number of iterations required for sparse recovery. The maximum number of iterations is set as* 50.

In this experiment, we set $n = 1000$ and $m = \beta n$, where the ratio $\beta = m/n$ is ranged from 0.1 to 0.6 with stepsize 0.025. For every such ratio, a random k-sparse vector x^* with $k = \lfloor m/10 \rfloor$ and 50 random matrices $A \in \mathbb{R}^{m\times n}$ were generated. We set $y := Ax^*$ as the measurements of x^* for every generated matrix A. The maximum number of iterations was set to be 50 for all algorithms. The average numbers of iterations required by the ROTP, ROTP2, and ROTP3 to meet the recovery criterion $\|x^p - x^*\|/\|x^*\|_2 \le 10^{-2}$ are summarized in Figure 5.1(b) which shows that the ROTP3 needs on average a smaller number of iterations than the ROPT2, and both need a smaller number of iterations than the ROTP to meet the recovery criterion. When the ratio is relatively high, all these algorithms only require a small number of iterations to meet the criterion. However, the average number of iterations required by these algorithms increases as the ratio m/n decreases. When the ratio m/n drops to a certain threshold, the number of iterations required by the ROTP to meet the recovery criterion goes above and beyond the prescribed maximum number of iterations, and thus the algorithm terminates after 50 iterations.

The other two experiments were carried out to compare our algorithms with several existing ones in terms of success frequencies of signal recovery. The first comparison was done for the k-sparse signal recovery with noisy measurements. The second comparison was done for both noisy signals and noisy measurements. We use the algorithms to recover, respectively, the sparse vectors $x^* \in \mathbb{R}^{1000}$ with different sparsity levels $\|x^*\|_0 \le 4k$, where $k = 25, 26, \ldots, 65$, and their noisy counterparts $\widetilde{x}$ which are approximately k-sparse. For every such sparsity level, we performed 50 random trials of the pair (A, x^*), where $A \in \mathbb{R}^{500\times 1000}$. In the first comparison, we set $y = Ax^* + \epsilon\theta$ as the measurements of x^*, where $\epsilon = 0.01$ and $\theta \in \mathbb{R}^n$ is a random noise vector with each component following a $\mathcal{N}(0, 1)$ distribution. We applied the IHT, HTP, ℓ_1-minimization, ROTP, ROTP2, and ROTP3 to these recovery problems, respectively, and we adopted $\|x^p - x^*\|/\|x^*\|_2 \le 10^{-2}$ as the stopping criterion. When an iterate x^p satisfies this criterion, the algorithm terminates and a "success" is counted; otherwise an "unsuccess" is counted. If the above criterion is not satisfied after the algorithm has been performed 50 iterations (which was set as the maximum number of iterations in our experiments), then the algorithm still terminates and an "unsuccess" is counted. In the second comparison, we generated A in the same way as the first comparison. The nonsparse vectors $\widetilde{x}$ were generated by adding the noises to the sparse



(a) Sparse signals and inaccurate measurements

(b) Noisy signals and inaccurate measurements

FIG. 5.2. *Comparison of the success frequencies of the algorithms for signal recovery with inaccurate measurements. For every sparsity level,* 50 *random examples were realized.*

vectors x^*, i.e., $\widetilde{x} = x^* + \widetilde{\epsilon}\widetilde{\theta}$, where $\widetilde{\epsilon} = 0.001$ and $\widetilde{\theta}$ is a random noise vector with each entry having a $\mathcal{N}(0,1)$ distribution. We then set $y := A\widetilde{x} + \epsilon\theta$ as the measurements of $\widetilde{x}$, where $\epsilon = 0.01$ and θ is a random noise vector with each entry following $\mathcal{N}(0,1)$. The stopping criterion for this case was chosen as $\|x^p - \widetilde{x}_S\|/\|\widetilde{x}_S\|_2 \le 10^{-2}$, where S is the index set for the $4k$ largest magnitudes of $\widetilde{x}$, where $k = 25, \ldots, 65$. The success rates of the algorithm are summarized in Figure 5.2, in which (a) is the result for the case in which y is inaccurate and x^* is k-sparse, and (b) is the result for both noisy measurements and noisy signals. The experiments indicate that the ROTP, ROTP2, and ROTP3 remarkably outperform the traditional IHT and HTP that fail to recover the vectors with sparsity in the abovementioned ranges. More interestingly, the ROTP2 and ROTP3 outperform the ROTP and remarkably outperform the ℓ_1-minimization method, especially in noise scenarios. The experiments indicate that the success rates of ℓ-minimization is somewhat sensitive to the noise level of the signals. Our algorithms, however, are more robust than ℓ-minimization for noisy signal recovery.

6. Conclusions and future work. The oscillation phenomenon in HTPs can be overcome by linking the k-thresholding with residual reductions. The optimal thresholding technique introduced in this paper naturally leads to the relaxed optimal k-thresholding pursuit (ROTP) and its enhanced counterparts, ROTP2 and ROTP3, which turn out to be efficient numerical methods for sparse optimization problems. The experiments indicate that the residual can be successively reduced in the course of iterations of the proposed algorithms, and thus the iterates generated by these algorithms move in a stable manner towards the solution of the sparse optimization problems. The essential idea for this new development is that *the hard thresholding operator should be applied to a compressible vector, instead of any vector.* The OT and OTP provide a fundamental basis for the development of such efficient numerical methods. Motivated by this study, several research directions are worthwhile to pursue in the near future. For instance, the recovery bound $\delta_{2k} \le \tau^*$ in Theorem 4.3 goes beyond the bounds for traditional hard thresholding methods. However, this bound remains largely theoretical from the perspective that directly solving the binary quadratic optimization problem in OT or OTP remains challenging, especially in high-dimensional settings. How to use the modern integer programming techniques

to deal with the subproblems in OT and OTP without relying on the convex relaxation technique is one question for interesting future work. In addition, the study in this paper demonstrates that the ROPT, ROPT2, and ROTP3 derived from convex relaxation are very efficient thresholding methods compared with existing ones. However, the first convergence result for the ROTP was shown in this paper under the condition $\delta_{3k} \leq 1/5$ which is relatively restrictive. Whether this result can be improved is also a worthwhile question to address in the near future. Moreover, the optimal thresholding technique introduced in this paper can be used to stabilize any sparsity-seeking procedures provided that the hard thresholding operator is involved in the procedure, such as compressed sampling matching pursuits, subspace pursuits, and the graded HTPs. So a further development for these procedures can be anticipated as well. We use this paper to develop a preliminary theory but a key step towards such a further development.

Acknowledgments. We would like to thank two anonymous reviewers and the Associate Editor for their helpful comments and suggestions which helped improve this paper.

REFERENCES

[1] A. Beck and Y. C. Eldar, *Sparse signal recovery from nonlinear measurements*, in Proceedings of the 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 5464–5468.

[2] A. Beck and Y. C. Eldar, *Sparsity constrained nonlinear optimization: Optimality conditions and algorithms*, SIAM J. Optim., 23 (2013), pp. 1480–1509.

[3] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[4] D. Bertsimas, A. King, and R. Mazumder, *Best subset selection via a modern optimization Lens*, Ann. Statist., 44 (2016), pp. 813–852.

[5] J. D. Blanchard, J. Tanner, and K. Wei, *CGIHT: Conjugate gradient iterative hard thresholding for compressed sensing and matrix completion*, IEEE Trans. Signal Process., 63 (2015), pp. 528–537.

[6] T. Blumensath, *Accelerated iterative hard thresholding*, Signal Process., 92 (2012), pp. 752–756.

[7] T. Blumensath and M. E. Davies, *Iterative hard thresholding for sparse approximation*, J. Fourier Anal. Appl., 14 (2008), pp. 629–654.

[8] T. Blumensath and M. E. Davies, *Iterative hard thresholding for compressed sensing*, Appl. Comput. Harmon. Anal., 27 (2009), pp. 265–274.

[9] T. Blumensath and M. E. Davies, *Normalized iterative hard thresholding: Guaranteed stability and performance*, IEEE J. Sel. Top. Signal Process., 4 (2010), pp. 298–309.

[10] J.-U. Bouchot, *A generalized class of hard thresholding algorithms for sparse signal recovery*, in Approximation Theory XIV: San Antonio 2013, Springer Proceedings in Mathematics & Statistics, G. Fasshauer and L. Schumaker, eds., 83 (2014), pp. 45–63.

[11] J.-U., Bouchot, S. Foucart, and P. Hitczenki, *Hard thresholding pursuit algorithms: Number of iterations*, Appl. Comput. Harmon. Anal., 41 (2016), pp. 412–435.

[12] A. M. Bruckstein, D. L. Donoho, and M. Elad, *From sparse solutions of systems of equations to sparse modeling of signals and images*, SIAM Rev., 51 (2009), pp. 34–81.

[13] C. Buchheim and E. Traversi, *Quadratic combinatorial optimization using separable underestimators*, INFORMS J. Comput., 30 (2018), pp. 424–637.

[14] E. J. Candès, *The restricted isometry property and its implications for compressed sensing*, C. R. Math. Acad. Sci. Paris, 346 (2008), pp. 589–592.

[15] E. J. Candès and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 51 (2005), pp. 4203–4215.

[16] E. J. Candès, M. Wakin, and S. Boyd, *Enhancing sparsity by reweighted ℓ_1 minimization*, J. Fourier Anal. Appl., 14 (2008), pp. 877–905.

[17] V. Cevher, *On accelerated hard thresholding methods for sparse approximation*, in Proceedings of SPIE Optical Engineering + Applications, Wavelets and Sparsity XIV, 2011, 813811.
[18] W. A. Chaovalitwongse, I. P. Androulakis, and P. M. Pardalos, *Quadratic integer programming: Complexity and equivalent forms*, in Encyclopedia of Optimization, C. Floudas and P. Pardalos, eds., Springer, Boston, MA, 2008.
[19] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput., 20 (1998), pp. 33–61.
[20] W. Dai and O. Milenkovic, *Subspace pursuit for compressive sensing signal reconstruction*, IEEE Trans. Inform. Theory, 55 (2009), pp. 2230–2249.
[21] I. Daubechies, M. Defries, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457.
[22] D. L. Donoho, *De-noising by soft-thresholdinng*, IEEE Trans. Inform. Theory, 41 (1995), pp. 613–627.
[23] D. L. Donoho and I. Johnstone, *Idea spatial adaptation via wavelet shrinkage*, Biomatrika, 81 (1994), pp. 425–455.
[24] M. Elad, *Why simple shrinkage is still relevant for redundant representation*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5559–5569.
[25] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, New York, NY, 2010.
[26] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, Cambridge, UK, 2012.
[27] M. Figueiredo and R. Nowak, *An EM algorithm for wavelet-based image restoration*, IEEE Trans. Image Process., 12 (2003), pp. 906–916.
[28] M. Fornasier and R. Rauhut, *Iterative thresholding algorithms*, Appl. Comput. Harmon. Anal., 25 (2008), pp. 187–208.
[29] S. Foucart, *Sparse recovery algorithms: Sufficient conditions in terms of restricted isometry constants*, in Approximation Theory XIII: San Antonio 2010, M. Neamtu and L. Schumaker, eds., Springer, New York, NY, 2012, pp. 65–77.
[30] S. Foucart, *Hard thresholding pursuit: An algorithm for compressive sensing*, SIAM J. Numer. Anal., 49 (2011), pp. 2543–2563.
[31] S. Foucart an M. Lai, *Sparsed solutions of underdetermined linear systems via ℓ_q-minimization for $0 \leq q \leq 1$*, Appl. Comput. Harmon. Anal., 26 (2009), pp. 395–407.
[32] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Springer, New York, NY, 2013.
[33] R. Garg and R. Khandekar, *Gradient descent with sparsification: An iterative algorithm for sparse recovery with restricted isometry property*, in Proceedings of the International Conference on Machine Learning, 2009, Montreal, Canada, pp. 337–344.
[34] M. Grant and S. Boyd, *CVX: MATLAB Software for Disciplined Convex Programming*, version 1.21, April 2017.
[35] K. Herrity, A. Gilbert, and J. Tropp, *Sparse approximation via iterative thresholding*, in Proceedings of the 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 624–627.
[36] L. Landweber, *An iteration formula for Freholm integral equations of the first kind*, Amer. J. Math., 73 (1951), pp. 615–624.
[37] K. Lange, *MM Optimization Algorithms*, SIAM, Philadelphia, PA, 2016.
[38] R. Khanna and A. Kyrillidis, *IHT Dies Hard: Provable Accelerated Iterative Hard Thresholding*, preprint, arXiv:1712.09379 [math.OC], 2017.
[39] N. Kingsbury and T. Reeves, *Redundant representation with complex wavelets: How to achieve sparsity*, in Proceedings of the 2003 IEEE International Conference on Image Processing, Barcelona, pp. 45–48.
[40] A. Kyrillidis and V. Cevher, *Matrix recipes for hard thresholding methods*, J. Math. Imag. Vision, 48 (2014), pp. 235–265.
[41] H. Liu, M.-C. Yue, A.M.-C. So, and W.-K., Ma, *A discrete first-order method for large-scale MIMO detection with provable guarantees*, in Proceedings of the IEEE 18th Internal Workshop on Signal Processing Advances in Wireless Communications, 2017.
[42] A. Maleki, *Coherence Analysis of Iterative Thresholding Algorithms*, in Proceedings of the Forty-Seventh Annual Allerton Conference on Communication, Control, and Computing, 2009, pp. 236–243.
[43] S. Mallat and Z. Zhang, *Matching pursuits with time-frequency dictionaries*, IEEE Trans. Signal Process., 41 (1993), pp. 3397–3415.

[44] S. Mallat, G. Davis, and Z. Zhang, *Adaptive time-frequency decompositions*, SPIE J. Opt. Eng., 33 (1994), pp. 2183–2191.
[45] A. Miller, *Subset Selection in Regression*, CRC Press, Boca Raton, FL, 2002.
[46] B. K. Natarajan, *Sparse approximate solutions to linear systems*, SIAM J. Comput., 24 (1995), pp. 227–234.
[47] D. Needell and J. A. Tropp, *CoSaMP: Iterative signal recovery from incomplete and inaccurate samples*, Appl. Comput. Harmon. Anal., 26 (2009), pp. 301–321.
[48] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Springer Science and Business Media, New York, NY, 2013.
[49] T. H. Reeves and N. G. Kingsbury, *Overcomplete image coding using iterative projection-based noise shaping*, in Proceedings of the 2002 IEEE International Conference on Image Processing, Rochester, pp. 597–600.
[50] J. Starck, M. Nguyen, and F. Murtagh, *Wavelet and curvelet for image deconvolution: A combined approach*, J. Signal Process., 83 (2003), pp. 2279–2283.
[51] S. Voronin and H. J. Woerdeman, *A new iterative firm-thresholding algorithms for inverse problems with sparsity constraints*, Appl. Comput. Harmonic Anal., 35 (2013), pp. 151–164.
[52] J. A. Tropp and A. C. Gilbert, *Signal recovery from random measurements via orthogonal mathcing pursuit*, IEEE Trans. Inform. Theory, 53 (2007), pp. 4655–4666.
[53] Y.-B. Zhao, *Sparse Optimization Theory and Methods*, CRC Press, Boca Raton, FL, 2018.
[54] Y.-B. Zhao and M. Kočvara, *A new computational method for the sparsest solutions to systems of linear equations*, SIAM J. Optim., 25 (2015), pp. 1110–1134.
[55] Y.-B. Zhao and Z.-Q. Luo, *Constructing new reweighted ℓ_1-algorithms for sparsest points of polyhedral sets*, Math. Oper. Res., 42 (2017), pp. 57–76.
[56] Y.-B. Zhao and D. Li, *Reweighted ℓ_1-minimization for sparse solutions to underdetermined linear systems*, SIAM J. Optim., 22 (2012), pp. 893–912.